FDI LOCATION AND SPILLOVER: EVIDENCE FROM HUNGARY

By
Gábor Békés

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
CENTRAL EUROPEAN UNIVERSITY
BUDAPEST, HUNGARY
FEBRUARY 2007

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS

The undersigned hereby certify that they have read and recommend to the Department of Economics for acceptance a thesis entitled "**FDI Location and Spillover: Evidence from Hungary**" by **Gábor Békés**

Dated: February 2007

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Principal Advisor: ______________________
László Halpern

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Advisor: ______________________
Gianmarco I.P. Ottaviano

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Reader: ______________________
Gábor Kézdi

Approved for the Department of Economics PhD Committee

PhD Program Director: ______________________
László Mátyás

CENTRAL EUROPEAN UNIVERSITY

Date: **February 2007**

Author: **Gábor Békés**

Title: **FDI Location and Spillover: Evidence from Hungary**

Department: **Economics**

Degree: **Ph.D.**

HEREBY I TESTIFY THAT THIS THESIS CONTAINS NO MATERIAL ACCEPTED FOR ANY OTHER DEGREE IN ANY OTHER INSTITUTION AND THAT IT CONTAINS NO MATERIAL PREVIOUSLY WRITTEN AND/OR PUBLISHED BY ANOTHER PERSON, EXCEPT WHERE APPROPRIATE ACKNOWLEDGEMENT IS MADE.

Signature of Author

Table of Contents

Abstract

This thesis is about determinants of FDI location as well as its impact on domestic firms. Following a brief introduction, this thesis consists of three papers that analyze the motives behind the location of foreign firms as well as some of the implications of their presence.

In a new economic geography framework with input-output linkages, Chapter Two analyses decisions made by foreign firms about their location within Hungary during the 1993-2002 period. Business to business contacts are modeled by creating several corporate customer and supplier access measures, and several access variables seem to affect location decisions. Chapter Two tests the effect of local development, regional and local policies on the location decisions of foreign owned firms in Hungary. It is found that lower local taxes, denser road network and local concessions may lure firms in. Chapter Three looks at the influence of foreign firms - once they are settled. The presence of multinational firms may be beneficial to domestic firms should superior productivity spill over to domestic firms. We argue that larger and more productive firms are more able to reap benefits of spillovers from multinationals firms.

Empirical investigations carried out in this thesis rely on a rich dataset with information on a census of firms in Hungary for the 1992-2003 period, and we focus on manufacturing firms.

Keywords: economic geography, industrial location, FDI, multinationals, regional policy, discrete choice models productivity.
JEL classification: C35, D21, F23, F14, H7, R3, R12

Acknowledgements

First of all, I would like to express gratitude towards three professors who have been instrumental in shaping this thesis. László Halpern was both my MA and PhD thesis supervisor. His MA course in international trade guided my studies towards trade, and his inclusion of applied work in the field must have been an impetus for me to study empirical applications. Second, I am grateful for LSE professor Anthony Venables whose class on economic geography back in 1999 spurred my interest in the subject. At that time, he started teaching draft of his seminal book written. Learning about agglomeration forces and market potential for the first time was an eye-opener experience, and the idea of non-linearity in development or bifurcations in agglomeration shaped the way I look at the world. Gianmarco Ottaviano has been my external PhD supervisor. I met Gianmarco at a seminar in Hamburg and he has been guiding my work ever since. His introduction of key references was quite important and his suggestion that one should consider his own comparative advantage in research seems a rather good advice.

The Central European University has always been helpful during my research project. Most importantly, the University allowed me to meet people who helped a great deal. I would like to thank László Mátyás for encouragement in early years, Gábor Kézdi for useful talks about methodology and substance and John Earle for beeing supportive both as professor of economics and head of the PhD program. I have benefitted a great deal from talking to Álmos Telegdy discussing data management and various research issues. I am also thankful for Carlo Altomonte, Thierry Mayer, Fabrice Defever, for reading earlier versions and providing valuable advice. Also, I was very happy to work with my PhD fellows, especially Gergely Csorba, Ádám Reiff, Gábor Futó and Ádám Szentpéteri, I learned a lot and enjoyed spending time and discussing problems with them. Further, the University supported several research trips via its Travel Grants and Doctoral Research Grant.

The Institute of Economics of the Hungarian Academy of Sciences has provided continuous support both in terms of personnel and data provision, thanks are due to Gábor Körösi. I thank Magyar Nemzeti Bank for having me as visiting researcher, as well as Zoltán Jakab, Péter Benczúr and István Kónya for providing useful comments.

The second chapter is based on a paper I started to write in 2003 when I was visiting the Magyar Nemzeti Bank whose hospitality is kindly acknowledged. Working with a large dataset such as the Hungarian corporate dataset was both a hard and happy task. The fourth chapter has relied on a great work by Gábor Kátay and Zoltán Wolf and I benefitted from discussions with them a great deal. The third chapter is based on joint research program with Balázs Muraközy (CEU, MTA KTI). It was a fun experience and fruitful collaboration that helped understand difficulties of using policy data.

I have also benefited from comments from participants at conferences and seminars where earlier versions of chapters were presented. I am especially grateful for HWWA Hamburg. I spent a useful two months visiting the University of Bologna. Under Gianmarco's supervision, several key ideas of chapter 4 emerged during my stay. I also benefited from two stays in Paris at University of Paris I / TEAM Institute. I was fortunate to work with Farid Toubal (Pantéon-Sorbonne Economics and PSE, University of Paris 1) and Jörn Kleinert (Eberhard-Karls-University of Tübingen) on a project that resulted in a coauthored article, presented in Chapter 4.

Needless to say, all remaining errors are mine.

***

Last, but in no way least, let me take the opportunity to thank my family for their patience and love. Their support has indeed been invaluable to this thesis. This work is in loving memory of my father who remained interested in my research topics till his last days.

Budapest
February, 2007

Gábor Békés

Chapter 1

Location of manufacturing FDI in Hungary: How important are inter-company relationships?

Abstract

In a new economic geography framework with input-output linkages, this study analyses decisions made by foreign firms about their location within Hungary during the 1993-2002 period. Business to business contacts are modeled by creating several corporate customer and supplier access measures using a rich corporate dataset of all Hungarian firms. Various econometric specifications of both discrete choice and count data models are applied. Results show that there are agglomeration forces at work and input-output linkages may be providing a key reason for co-location. For example, access to firms operating in the same industry as the new firm and a proximity of potential customers throughout the country are shown to be persistently important determinant of location choice. Distinctions between early and later years as well as foreign acquisition (such as privatization) and greenfield FDI are made, and a few difference between decisive factors are shown.

1.1 Introduction and motivation

In Central and Eastern Europe, rapid changes and restructuring in manufacturing have taken place since 1990 and spatial inequality has risen substantially along with new investments by firms entering countries previously closed to foreigners. This period may be characterized by economic liberalization and opening up markets to foreigners. Prior to 1989, there has been hardly any foreign direct investment in Hungary, while by 2003, the stock of FDI reached almost 50% of the GDP. One might expect that in the lack of specific policies, capital would spread out rather evenly in a small country like Hungary.

However, this has not been the case. Instead, agglomeration of new firms and a further spatial polarization have been visible phenomena in general and in most sectors of the industry. For example, the GDP per capita ratio of the richest and poorest two counties rose from 2.3 in 1993 to 2.9 in 2000. Considering all nineteen (without the capital), roughly similarly-sized counties in Hungary, our data suggest that foreign entries were often concentrated in a few regions. Table 1. reports some basic features of regional properties suggesting a great deal of variation in terms of FDI intensity. While Budapest attracted 120 new firms per 100.000 inhabitants while the same measure is less than 20 for the worst performing two counties.[1]

Looking at dynamics, new foreign entry has not contributed in any way to a convergence of income. As shown in Table 2, in terms of new FDI projects, the share of Budapest has been rather stable, and so was the share of the three richest (per capita GDP) and poorest counties, with only a very slight convergence at the last two years.

One reason behind observing the agglomeration tendency in Hungary may have been that "new firms have a high propensity to settle at places where economic activities are already established"[2]. With this proposition in mind, we look for an explanation of the increasing spatial difference by analyzing a narrow location choice problem, that of new foreign firms in manufacturing. This problem has been treated both in an international and a national setup. As for the case of several countries, Head & Mayer (2005) look at Japanese investments carried out in the European Union and find that market potential, agglomeration variables as well as several traditional explanatory variables (e.g. taxes) determine how firms pick a region. Taking one country, Barrios et al. (2003) look at multinationals' location choice in Ireland and find that agglomeration forces contributed substantially to location choices and proximity

[1] Interestingly, the same pattern of agglomeration is true for a set of countries in the region: manufacturing of electronic devices by firms such as Flextronics in Central and Eastern Europe 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. (For details see Barta (2003).)

[2] (Ottaviano et al. 2002, p.7.)

to major ports and airports was also helpful. Cieslik (2003) uses a Poisson model on 50 Polish regions to show that proximity of key export targets, industry and service agglomeration, and road network are the important magnets for foreign investment. In some respect, the Amiti & Javorcik (2003) study, which considers subsidiaries of multinational firms in China, is the closest to this paper as we also try to disentangle effects of proximity to suppliers and customers.

As for Hungary, Boudier-Bensebaa (2005) focuses on the agglomeration effect and estimates motives of location choice using regionally aggregated data. The distribution of the FDI stocks in counties are found to be related to labour conditions and manufacturing density and the number of existing enterprises.

In this paper, we analyse decisions made by foreign firms about their location within Hungary in a new economic geography framework with input-output linkages. This work aims at making three contributions. First, the empirical investigation in this paper is related to a multi-industry model of new economic geography with an emphasis on input-output linkages. As a result, the focus of this paper is to test the significance of some key variables stemming from such a framework. Firm-to-firm contacts are modelled by creating several corporate customer and supplier access measures for firms based on firm-level data. Results showed that there are agglomeration forces at work and input-output linkages work their way through supplier and market access providing a key reason for co-location. For example, access to firms operating in the same industry as the new firm as well as proximity of potential customers throughout the country seemed to be a persistently important determinant of location choice. To ensure robustness of our results, various econometric specifications of both discrete choice and count data models are applied and compared.

Second, unlike in several preceding work, all access variables are generated from firm-level data of corporate balance sheets and income statements. The dataset includes all firms in manufacturing following firms throughout their lifetime -during the sample period of 1992-2002. Having determined an entry year after 1992, final corporate decisions may be looked at instead of announcement of investment projects that may or may not have been realised. Wages are generated from annual labour market surveys that allowed to create industry and for most cases, job-type specific wages helping detect the discouraging impact of high wages.

Third, rather than looking at a large and developed country, investigation is carried out on a small European economy that has just gone through economic transition involving almost unprecedented rapid market liberalisation. With hardly any capital inflow before 1990, our data for 1992-2002 allows to track the entry of most foreign firms operating in the economy. Distinction between early and late years can be made, and privatisation (or in broad terms: foreign acquisition) may be separately treated from greenfield FDI. We found that high wages seems to have been more of a deterrent in the early nineties and foreigners picked privatized companies on similar grounds than did domestic enterprises.

The paper is organised as follows. Section two summarizes the theoretical background of location choice. Datasets and variables are described in section three discussing the creation of the access variables as well. Section four presents econometric methodologies along with all the results and robustness checks. The last section concludes and talks about potential advances. Details on the nature of our firm-level data as well as description of how data problems were handled are relegated to the appendix.

1.2 Theoretical background: input-output linkages and geography

Why has agglomeration of new FDI projects been such a visible feature of development? By helping understand the nature of differences between developed and underdeveloped regions, the recent strand of new economic geography (or NEG) aims to shed light on this phenomenon.[3] In many NEG models built to match US or British data, labour migration is essential for agglomeration forces to work: an increased population generates greater demand inviting more firms to settle in a larger city, and this allows for a lower import bill and hence, lower living costs in general. However, even in the long run, labour migration is rather low in continental Europe. Thus, another agglomeration force is required to explain the desire to co-locate in spite of low migration propensities. One possible solution is the incorporation of input-output (I-O) linkages that explicitly capture trading costs between firms. Indeed, the presence of inter-company sales, modeled for the NEG literature by Venables (1996) and Krugman & Venables (1995), can provide this motivation and hence, is an important

[3]For details, see for example Baldwin et al. (2003) or Ottaviano et al. (2002). An excellent survey of key hypotheses emerging from models of new economic geography and their mixed empirical support can be found in Head & Mayer (2004).

subject for empirical work.

The theoretical framework for our analysis aims to emphasise business to business relations as a key driver of location decisions. 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. In the background of the empirical investigation to come, stands a multi-country and multi-industry model following (Fujita et al. 1999, Chapter 15A) using the "classic" ingredients of new economic geography based on the monopolistic competition of Dixit & Stiglitz (1977). One key aspect of firm-to-firm relationship here is related to input-output linkages that were introduced by Krugman & Venables (1995) in order to model the fact that firms sell goods not only to consumers but to other firms as well.

In the framework, there is monopolistic competition in all sectors producing a range of differentiated goods.[4] The paper focuses on manufacturing. The agriculture sector, which has been present in many similar models, will be overlooked. True, a dispersion force will be lost but in the lack of large-scale migration, wages and local consumer demand should be strong enough to foster agglomeration.

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. Profit for each firm depends on firm- and industry-level characteristics. Firm-level characteristics such as technology advantage over industry peers and quality of management are unobserved. However, these features are assumed to be independent from the choice of location. Another determinant of a given firm's profit depends on such industry features as (average) technology, skill requirements, transaction costs and location of markets. These are indeed region-dependent factors. Profit for firm i in industry j and region r will come from these two terms, but assuming that they are separable allows to disregard firm-specific determinants. We assume 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. Thus, profit is simply the difference between revenues and costs (i.e. f.o.b. sales prices less marginal costs, times quantity): $\Pi_r^j = (p_r^j x_r^j - mc_r^j x_r^j)$.

The marginal cost function of a representative firm in industry j and region r is

[4] A more detailed description of the setup may be found in the working paper version, Békés (2005)

$$mc_r^j = wage_r^{a_j}(GP_r^j)^{\mu_j}(\mathbf{b}_r^j)^{\delta_j} \tag{1.2.1}$$

as the marginal cost depends on $wage_r$, the nominal wage, GP_r^j the composite index of intermediate good prices (P_r^j) and a $\mathbf{b_r^j}$ vector of other location dependent non-labor factors of the locally consumed production such as communication infrastructure.

Firms use a set of goods produced by firms in other industries that are aggregated by a CES subutility function into a composite good. The intermediate good price index, defined as 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 a greater quantity of necessary intermediate goods is 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 a weighted average (by the number and size of relevant firms) of f.o.b. prices that already include an iceberg type transport cost, $\tau_{l\_r}^j \geq 1$: (i.e. for the home region only $\tau_{l\_l}^j = 1$).

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. The composite price index of intermediate goods then takes into account the source of various goods, using the national input-output table to determine input shares. Note that the way demand is set up creates a backward linkage: firms want to be close to their markets and potential customers.

Unlike in Fujita et al. (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 the dynamic effects (e.g. on the labour market) of an investment. In the long run equilibrium, prices are adjusted taking externalities into account. For example, wages or land prices will reflect benefits of agglomeration and lower prices in one region will only signal poorer circumstances. In the short run, disequilibria may exist and entry of firms (bidding up wages and input prices) shall be considered as a force to bring prices closer to their equilibrium value. The main goal of this exercise is to obtain a corporate profit function that will be linked to the settlement decisions of firms in the empirical work. Indeed, the profit function must capture both key notions introduced in the previous section. First, access to markets is incorporated both for firms and for final consumers via demand and price variables. Second, agglomeration economies will be captured by some $\mathbf{b}_r^j$ variables as well as some of the access variables.

Demand for products by a firm operating in industry j and in region r comes from three sources: final consumers, corporate customers and agents abroad. Note that the way demand is set up allows the introduction of some of the key business to business relationships as it creates a backward linkage: firms want to be close to their markets and potential customers. In the Fujita et al. (1999) framework, it can be shown that aggregate demand, AD_r^j is a weighted sum of these types of buyers.

$$AD_r^j := \left[\sum_{l=1}^{R} (\frac{P_l^j}{\tau_{l\_r}^j})^{\sigma_j - 1} \left(\mu_l^j INC_l + \sum_{k=1}^{K} io^{jk} X_l^k \right) + \left[(\tau_{x\_r}^j)^{-1} FAD_r^j \right] \right] \tag{1.2.2}$$

The first term denotes the price of a good as a function of both its factory price and the transport costs. The second one is the sum of local final demand and corporate demand. Final good demand is proxied by purchasing power of consumers that is measured by the variable INC, which is decomposed into the number of inhabitants and income per capita. Business demand is measured by industry level sales weighted by required input shares. Note that market access to corporate customers will be denoted as MA_r^j. The third term represent foreign total demand or, FMA_r^j.

This structure generates a profit function that is determined by labour costs, aggregate demand, intermediate good prices, and other cost factors. Taking logs, we get a linear relationship.

$$\ln \Pi_r^j = a_j \ln (wage_r) + \mu_j \ln(GP_r^j) + \delta_j \ln(\mathbf{b}_r^j) + \left[AD_r^j \right] \tag{1.2.3}$$

Unlike in models of international geography such as Head & Mayer (2005), the intermediate good price index (GP_r^j) cannot be measured directly owing the lack of regional price data.

Instead, we assume that intermediate good prices primarily depend on the proximity of suppliers; and given the market structure, the intermediate good price index will be negatively correlated with an access measure to producers of such goods. Note that this supply access will be measured similarly to the market access using input requirement (rather than output) coefficients as weights.[5]

[5]This is of course a similar idea used in Redding & Venables (2004). There are other ways to

As a control variable, we include a regional concentration index for the given industry control for market structure. It is expected that a less concentrated sector in a region allows for lower prices (or greater variety). Concentration will be measured by the region, industry specific Hirschman-Herfindahl index.

Another option could have been the use of another proxy of intermediate good prices such as energy. However, in Hungary, energy prices have been nationally regulated during the 1992-2003 period, and no regional price disparity has occurred as a result of price liberalization measures.[6]

The vector of cost factors ($\mathbf{b_{r(t)}}$) includes some basic features of development that are not industry specific. A more developed county should yield lower transaction costs and hence, marginal cost of production. We use several such measures and look for a positive relationship between development and location choice. As for the labour market, $wage_{r(t)}$ measures the local wage. Wage variables were calculated from the LMS data and reflect (gross) labour costs that should be expected by a firm looking to settle in the given county.

Of course, firms pay taxes and receive investment support. However, in Hungary, local economic policy is not defined by counties but determined at the settlement level, and regional tax incentives are relative novelty, so it was assumed that region specific state intervention is zero.

Finally, we need to introduce the time dimension that has been so far overlooked. The background model, in the lack of specific dynamics, gives no guidance about the timing. In the literature explanatory variables are sometimes considered at time t as well as $t-1$. In this work, 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). Hence, firms consider information available at $t-1$, decide and start building that becomes a firm in time t. The econometric support stems from a requirement to try to avoid endogeneity, and lagging will free the model of simultaneity bias (lagging makes it unnecessary to substract the actual firm's output of composite output (access) variables). We also need to assume that firms at time t

estimate supplier access. Head & Mayer (2006) follow Redding & Venables (2004) and use bilateral trade data to uncover coefficient of what they call the real market potential that has the supply access in its denominator. In our case, trade data between counties simply do not exist making such step impossible.

[6]On a narrow sample, petrol prices were also checked and no disparity at the county level was found despite sizeable differences by the city and exact location.

considering values of explanatory variables at time $t-1$, pick a county independently of each other. Agglomeration works as firms locate close to other firms that had settled previously, but there is *no* strategic interaction between firms settling at time t. This is a necessary assumption for using the logit model.

As a result, our expected profit function for a firm i depends on :

$$\begin{aligned}\pi^j_{r(t)}(i) = \alpha_1 wage^j_{r(t-1)} + \alpha_2 INC_{r(t-1)} &+ \beta_1 SA^j_{r(t-1)} + \beta_2 MA^j_{r(t-1)} + \beta_3 FMA^j_{r(t-1)} + \\ &+ \gamma b^j_{r(t-1)} + \zeta^j_{r(t)}(i) \end{aligned} \qquad (1.2.4)$$

where the error term, $\zeta^j_{r(t)}(i)$ includes all the non-observed variables.

1.3 Description of the data and variables

For this study, several sources of data are used. Most importantly, our corporate dataset to be described in this section includes all Hungarian manufacturing firms for eleven years. Wage data are gathered from annual labour surveys, while most control variables are produced by the Central Statistics Office (KSH) in Hungary. Table 5 reports basic data on all our variables.

1.3.1 The corporate dataset

There are two key datasets in the study. The corporate dataset used here, is based on annual balance sheet data submitted to the Hungarian Tax Authority (APEH) and is maintained by the Magyar Nemzeti Bank. The APEH dataset contains information on *all* registered, double entry book-keeping firms. The raw dataset provided by APEH was cleaned a great deal by the Magyar Nemzeti Bank as well as the CEU labour Project and several corrections and quality improvements were made the specific purpose of this project. (For details on the data cleaning, see the Appendix.)

Data include industry code, size of employment, share of foreign ownership and a county code. Data are available annually for the 1992-2002 period; entry dates may be generated as of 1993. The number of corporations in all sectors of the economy varies year to year, rising from 58 thousand in 1992 to 185 thousand in 2002. In manufacturing, there are 12 thousand firms in 1992 rising to 28 thousand in 2002.

The year of *firm birth* equals the year of first appearance in the dataset, i.e. the first year of submitting a report to the Tax Authority. For this is compulsory, there should be little error in measuring the entry date. *Foreign ownership* is defined whenever the foreign share in equity capital passes a 10% threshold. For foreign companies defined this way, the average foreign share is very high and results are quite robust to raising the threshold to 25%. Also, whenever foreign ownership is low at the beginning, in most cases it will rise substantially, often to 100%, after the first few years.

Overall, the dataset is composed of 5350 location settlements by firms with foreign ownership carried out in manufacturing. Only 4557 can be certainly considered as new investment rather than foreign acquisition, and this paper will be mostly dealing with these new investments only. Industries are grouped in sectors according to two-digit NACE codes. With merging some industries (e.g. clothing and leather), and excluding food production, there remain 15 sectors; Table 3. reports the main characteristics.

One problem may be the use of multi-plant firms where we model one location (its principal address) instead of the several plants used for production. Large multinationals are the main culprit. Fortunately, multinational companies mostly operate several firms in Hungary. For example, Dutch giant, Philips owns nine entities in four cities in Hungary, each being mainly responsible for one type of activity such as audio devices production, electric equipment manufacturing or retail (for more detail, see the Appendix.)

There are some coefficients that are not estimated but taken from other sources: Input-output table comes from Hungarian Statistics Office's publication on 1998 data (Hivatal (2001)). This is the only I-O table available for the time period used. However, the assumption that input requirements per sector 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.

1.3.2 From firm data to access variables

In this section, creation of variables, which are used in estimations, is explained.

First, unit transport cost ($\tau^j_{l\_p}$) is estimated by assuming a very simple relationship:

$$\tau^j_{l\_p} = dist_{l\_p} * V^j$$

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.

In reality we know little about coefficients of the relationship above. 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. Here, it is assumed that shipping a good to 200km costs twice the amount it does for 100km. Note, that this is higher than some estimates for international shipment costs (e.g. Hummels (2000)). However, our variable includes all costs related to doing business.

The value of a typical package of industrial output $V^j = (\$/kg)^j$ on 1km comes from the World Bank database on international freight costs. True, these figures are based on more developed market data, and aggregation will mask many features. However, it helps correct for the fact that it is cheaper to ship 100 euros worth of laptop PC than the same value of steel. (See Table 4.)

There are various ways to measure distance between counties ($dist_{l\_p}$), and here a simple method is chosen. First, using the KSH "T-STAR" database on settlements, the most important city per county is picked (i.e. with the largest number of manufacturing firms). Note that picking the key city was straightforward for all but one case, the largest city was at least twice the size of the second. Second, distance between any two counties is defined by measuring the road distance between the representative cities. It is assumed that goods are transported by trucks only, and that vehicles move at the same speed and costs are indifferent to road quality.

In its tax report, each company reports a sales figure that can be picked up from its balance sheet attached to the earnings report. Sales data for a firm i operating in industry j registered in region r at time t is denoted by: $x^j_{r(t)}(i)$. All access variables to be tested in forthcoming subsections are based on output figures per county and sector ($Y^j_{r(t)}$). These numbers are determined by aggregating sales figures from the balance sheet data for all the relevant firms i (in industry j and region r, time t): $Y^j_r = \sum_i x^j_{r(t)}(i)$.

Corporate access variables measure proximity to firms that may be relevant for a new company, and the access variable is the sum of output by firms weighted by distance and share in inter-company trade. From theory, we need one variable to measure demand (MA_r^j) and another one to proxy supplier access (SA_r^j). Bear in mind, that although supplier and market access variables are compiled in a similar fashion, they measure different types of variables. The market access is about demand, while the supplier access is just a proxy to (intermediate goods) prices.

Here, both variables are divided into two components: one to pick up access to local (internal or within county) firms and another one for non-local (external or outside the county) firms. We denote internal variables by an "loc", external ones by an "nat" suffix.

$$MA_r^j = \lambda_1 MAloc_r^j + \lambda_2 MAnat_r^j \tag{1.3.1}$$

The reason for such dichotomy comes from the suspicion that the effect of the relationship between firms is not linear to distance. For example firms clustered in one city or in a few cities close to each other, enjoy special agglomeration effects. Thus, separating local (intra-county) and non-local effects may be an easy way to model non-linearity while simplicity of calculation is kept. [7]

Theoretically these are the basic access variables we need. However, there may be (and as we will see it, there is indeed) a strong correlation between $SAloc$ and $MAloc$, and so is between $SAnat$ and $MAnat$. One possible reason for correlation between access variables is the fact that own industry output influences both the supplier and the market access variable strongly. This stems from the structure of commerce between firms: companies trade the most with other companies in the very same industry.[8] On average, intra-industry trade amounts to one third of total inter-company sales, and this exacerbates correlation between the $MAloc$ and $SAloc$ variables.

To remedy this, a new variable, $IPloc$ is introduced that measures own industry output only. This own industry access variable is essential in the analysis as it picks up several aspects of intra-industry transactions: trade, competiton and cooperation. In a way, it shows the strength of industrial clustering. Note that one further reason that makes worth locating close to one another is the potential of knowledge spillover. (See

[7] In a somewhat similar setup, Amiti & Javorcik (2003) created such aggregate access variables.

[8] This feature makes the use of models with two sectors, such as upstream and downstream industries, impossible.

Chapter 4 on this issue, where this variable is similar in approach to the horizontal spillover.)

Accordingly, corporate demand may be proxied by a local and a national (all regions except for the local one) industry dependent market access variables (local: $MAloc_r^j$, national: $MAnat_r^j$). Coefficients io^{kj} are used to give the share of intermediate goods produced by industry j purchased by other industries k.

$$MAloc_r^j = \sum_k^J io^{kj}(Y_r^k) \quad ; \quad MAnat_r^j = \sum_k^J io^{kj}(\sum_{l \neq r}^R \frac{Y_l^k}{\tau_{l\_r}^j}) \qquad (1.3.2)$$

$$SAloc_r^j = \sum_k^J io^{jk}(Y_r^k) \quad ; \quad SAnat_r^j = \sum_k^J io^{jk}(\sum_{l \neq r}^R \frac{Y_l^k}{\tau_{l\_r}^j}) \qquad (1.3.3)$$

$$IPloc_r^j = Y_r^j \quad ; \quad IPnat_r^j = (\sum_{l \neq r}^R \frac{Y_l^k}{\tau_{l\_r}^j}) \qquad (1.3.4)$$

The intermediate good price index is proxied by a supplier access variables, created in the same fashion, but eventually using rows instead of columns (and columns instead of rows) of the input-output table (io^{jk} instead of io^{kj}) that describes input requirements. Importantly, for both supplier and market access, we limit the input-output coefficients such that $i \neq j$. For cases when $k = j$, $IPloc_r^j$ and $IPnat_r^j$ were introduced.

Access to foreign markets influencing both demand and intermediate good prices, is measured by a single foreign access variable (FMA_r^j). This takes into account that export is a crucial determinant of the revenue of Hungarian firms and the average import share (by 1998) reached one-third for manufacturing. By the theory, the direct market access to foreign (i.e. in countries $n = 1, 2...N$) firms and customers should be taken into account.[9] However, due to data limitation problems, this paper proxies access to foreign markets by taking into account the distance to the key export borders.

$$FMA_r^j = \sum_{n \neq r}^N \frac{INC_n}{1 + \tau_{n\_r}^j} + \sum_i^{J+K} io^{ji} \left(\sum_{n \neq r}^N \frac{Y_n^j}{1 + \tau_{n\_r}^j} \right) \approx \sum_{n \neq r}^{N=4} \frac{ts_n}{1 + \tau_{n\_r}^j} \qquad (1.3.5)$$

[9] Amiti & 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.

where ts_n is the share of trade to the $n-th$ direction. [10]

Business access or BA_r^j picks up access to services such as banking, accounting or lodging, as a special determinant of production costs. For services are likely to be used locally, we only consider access to local business services.

$$BA_r^j = BAint_r^j = \sum_k^K io^{jk} Y_r^k \tag{1.3.6}$$

where k includes various service sectors of the economy.

1.3.3 Wages and other variables

As for the labor market, we assume that firms are too small to influence wages, and we have no forward looking wage-setting.

The cost of labour may be a crucial dispersion force and hence, its careful modeling is important. To get detailed wage data, a large employer-employee dataset is used that comes from annual Labour Market Survey (LMS) data compiled by the Ministry of Labour, containing employment data on a sample of some 140.000 employees per year. Employees are picked independently of their employers, i.e. we do not know exact wages of firms but instead wages that should be expected by *types* of companies. The large sample size allows for annual coverage of all industries in almost all counties.

This dataset allows to generate county level average wages for every year and county. Using the same annual labour survey, another labour cost variable is generated by averaging wages of employees in a given county as well as industrial sector. Out of the 3000 possible industry-county-year combinations, we were able to create 2737 industry specific wages directly from the data, while estimated the remaining 263 wage figures. Note that for such industry-year-county combinations, hardly any FDI investment has taken place - they were mostly used as counterfactuals for running the logistic regressions. Average regional and industry specific wages were created by weighting firm-level (gross) wage information by the size of firms (employment) so as to generate a wage level, a firm may expect when choosing a location. For every employee there is a description of the job, and this allows to create a special blue-collar wage for (almost) every industry and county.

[10] We used distance to the borders: West/Austria: Hegyeshalom; South/Croatia: Letenye; North/Slovakia: Komárom, East/Ukraine: Záhony ,Airport: Ferihegy/Budapest.

In this paper a few measures of development are chosen to take care of major sector independent variables[11]. Data come from the regional database of the Central Statistics Office.

The quality of local transport network is measured by the size of national road network within a given county, i.e. the size (in kilometers) of all national road (including motorways) divided by the area of the county (in km^2). Note that there has been little change in the size of the network throughout the observed period, the total size rose by about 3%. Another indicator is the provision of communication proxied by the number of landline telephone stations per county. This is another frequently used variable to proxy development of the infrastructure and thus, non-transportation linked transaction costs. Note however, that as a result of widespread use of mobile phones, this measure may have just turned to be a poor proxy by now. The number of students enrolled in higher education at institutions within the given county should proxy the abundance of management and R&D knowledge in the county.

In addition to measures of development, population density i.e. the size of population divided by size of the area of the county shall pick up an agglomeration externality: it may be cheaper to sell products when people are close to each other. However, a negative sign would suggest that this urbanization effect is outweighed by higher land prices.

1.4 Estimation methods and results

First, conditional logit (CL) models will be estimated to study the influence of input-output linkages, labour market conditions and market access on investment decisions in Hungary. A key achievement 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 the scarcity of information and errors made by analysts, the maximisation procedure per se is less than perfect. Thus, profit (or utility for consumers) is a random function of explanatory variables.[12]

[11]In a related paper (Békés & Muraközy (2005)) several more variables of local development and municipal policy are tested.

[12]For details, see Maddala (1983), (Train 2003, Chapter 3.)

1.4.1 Conditional logit model

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 (3) subject to uncertainty. Apart from the observed characteristics of firms, such as 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[13]. The explained variable is the location choice of firms so the choice variable is 1 if the investment took place in that particular county and 0 for the remaining 19 counties.

Taking all potential effects into account, a firm i (where $i \in \{1, ..., N\}$) of sector j (where $j \in \{1, ..., J\}$) locating in region r (where $r \in \{1, ..., R\}$) will attain a profit level dependent on various industry and region dependent variables. Importantly, not all of these variables matter, as the choice of region is independent on individual firm or industry characteristics. Thus, 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 profit equation (3) in parsimonious form for a firm i in industry j and region r is:

$$\pi^j_{r(t)}(i) = \gamma' b_{r(t)} + \lambda' d^j_{r(t)} + \varepsilon^j_{r(t)}(i) \quad (1.4.1)$$

In order to be able to use results of McFadden (1974), we need to assume that the error term, $\varepsilon^j_{r(t)}(i)$, is independently distributed across r and i, and has a type I extreme value (or Gumbel) distribution. The error term reflects unobserved terms as well as those that depend on individual firms. A crucial assumption is that unobserved characteristics do not cause correlation, i.e. errors are independent of each other. In other words, independence here requires that the error for one alternative provides no additional information about the error for another one. It is likely that this assumption would not hold very well for the data but the generality of the CL model allows for a detailed investigation. (For details and some remedies, see section 1.4.3 .)

[13] In the corporate database there are of course values for profit. However, for multinational companies they are heavily distorted by transfer pricing as well as various grants and incentives.

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

$$prob[\pi_{ij}(r) < \pi_{ij}(l)] = prob[\varepsilon_{irj} < \varepsilon_{ilj} + A_r - A_l + \gamma' b_r + \lambda' d_r^j - \gamma' b_l - \lambda' d_l^j] \quad (1.4.2)$$

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

$$P_r^j = P_{r|j} = \frac{\exp(\gamma' b_r + \lambda' d_r^j)}{\sum_{l=1}^{R} \exp(\gamma' b_l + \lambda' d_l^j)} \quad (1.4.3)$$

Estimation is carried out by maximising the log-likelihood:

$$\log L = \sum_{j=1}^{J} \sum_{r=1}^{R} n_r^j \log P_r^j$$

where n_r^j denotes the number of investments carried out in sector j of region r.

In most specifications, fixed effects are added to pick up possible level shifts caused by some omitted variables such as economic policy. As a result, equation (1.4.1) would become:

$$\pi_{r(t)}^j(i) = \delta_r + \gamma' b_{r(t)} + \lambda' d_{r(t)}^j + \varepsilon_{r(t)}^j(i) \quad (1.4.4)$$

where δ_r are location specific dummy variables. County level as well as NUTS2 region level dummies are introduced to the key equations.

Note that coefficients are approximations of the elasticity of the probability of choosing a particular county for the average investor.[14] For example, considering the most basic setup of specifications, a 10% increase in the local own industry access variable (or 10% rise in the output of the average firm or a 10% increase in the number of firms) would raise the probability of choosing that county by 2%.

[14]It can be shown that true coefficients are (1-p*) times the estimated figures, where p* is the average probability of choosing a region. Here, p*=1/20=0.05. Remember, that figures must be taken with care for the logit estimation is carried out with a normalization of the variance of the error term.

1.4.2 Results with the conditional logit model

Results with conditional logit are reported in Table 7. In order to control for unobserved county differences such as those stemming from first nature geography (such as rivers or hills), county or region dummies (choice specific constants) were introduced. To control for the specific case of Budapest, a capital dummy was added to the regional fixed effects. This had little effect on corporate access variables confirming the robustness of results. However, many explanatory variables, which depend upon location only, change little over time, and thus, would loose significance in due course.

By the basic specification (CL1), demand variables such as per capita income and size of population enter with the expected positive sign, while higher wages are associated with a lower likelihood of firm location.

The access to own industry output is strongly significant, i.e. the probability of a new foreign-owned firm choosing a county increases in the output of its very sector in that county. A 10% rise in output (average output per firm or number of firms) leads to a 2.2% gain in probability. The local own industry access remain very robust through any possible specification and model.

Positively significant are the national (external) market access variable as well as the local (internal) supplier access. Access to business services is also seem to matter for firms. These suggest that input-output linkages are important determinants of location choice.[15] Overall, the local presence of own industry is one of the most robust determinants of firm location: carmakers will try to settle where other firms in the motor vehicle industry are settled. Local suppliers outside one's industry matter but seem less important. The result that the national market access is always positive and significant suggests that firms would want to settle close to non-own industry customers, i.e. a steelmaker will consider all potential corporate customers when deciding about location.

At the same time, local market access and national supplier access enter with a negative sign, and so does national own industry output. Thus, market crwoding/competition effects are stronger for these variables. Other specifications confirm these results. These results contradict theoretical predictions but there may be several explanations for such result. Note however, when access variables were simply

[15]There are several descriptive evidence that suggest that supplier contacts are known to have been an important factor in the region.For example, in Hungary, suppliers to the Suzuki car plant are mostly settled within a close proximity to Suzuki, often in the same county.

aggregated into a local and a national corporate access variable, both entered with a significantly positive sign suggesting that overall, input-output linkages outweighed market crowding.[16]

First, the correlation between supplier and market access variables is strong (see Table 9) and this may have remained a problem despite previous efforts.[17] Note that this is not a unique problem of this study, several previous empirical works with both supplier and market access variables faced this correlation issue (Redding & Venables (2004)). In any case, multicollinearity is mostly a small sample issue, and we believe that our dataset is large enough to be able to disregard it. Further, when dropping one variable, the sign of others would not change. Second, these variables may pick up the impact of some negative externality of firm presence. A prime suspect is competition or market structure in general, which disregarded due to the assumption of infinite number of firms and the fixed markup result of monopolistic competition type NEG models.

For example, the negative sign on the access to national (external) output of the own industry, seems to suggest some sort of a competition effect outweighing any agglomeration effect. A negative sign may indicate such dispersion force: i.e. it is good to have similar firms close, but presence of too many firms in the neighborhood leads to market competition and under monopolistic competition, more varieties imply lower profits. At the same token, variables that enter with a positive sign may capture some other forces. Indeed, industries like to cluster for other reasons than input-output linkages as it is supported by the strong significance of the own industry access variable of the actual sector. One must remember that it is impossible to separate the key motives, such as labour pooling, knowledge spillover or a decrease in business costs due to information sharing. Despite our effort to filter out co-location due to supplier linkages, these problems remain important. Thus, results suggest that within a small area such as the county, agglomeration and input-output linkages are more important as captured by a strongly positive local own industry access variable. Market crowding outweighs these positive externalities for other counties in proximity as suggested by the strong negative sign of the national own industry access variable.

[16]Further results are available from the author on request.

[17]One potential reason for such result may be non-linearity in the data. To see this, I first looked at the access variables (in logs) and found that their distribution has a one-peaked distribution that looks not very different from a lognormal distribution. Second, quadratic terms were included to capture some sort of a hidden effect. It turned out that some quadratic variables were significant but they had no influence on any other variable.

Previous empirical work suggests that 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 a negative sign would suggest that high wages deter firm location. However, the empirical evidence is mixed with a slight leaning towards the opposite sign.[18] In the basic specification (CL1), the labour cost variable enters significantly with the expected negative sign confirming predictions of the theory. However, for other cases, the variable looses its significance and often, its sign changes.

There may be several explanations, this paper underlines just two such reasons. First, various industries use different types of labour in terms of skills, and hence, the industry mix of a region may or may not influence the aggregate wage variable. Second, individual industries use different types of labour in different shares. The share of blue-collar workers may vary a great deal among sectors and their wage may differ greatly depending on how skilled they are. Thus, in econometric models like those of this paper, wages may well reflect an "industry bias" as well as a "skill 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.

All specifications using industry specific average wages and blue-collar labour costs suggest that both variables enter significantly. A negative sign of the industry specific wage confirms the theoretically negative effect of high wage costs, while the positive sign of the blue-collar wage points to the notion that the skill bias is important for blue-collar workers.

Some specifications include non-industry dependent variables of $\mathbf{b_{r(t)}}$ such as the size of telephone or road network, both being positively related to firm location. This confirms generally held views that better infrastructure is key to attract FDI. The agglomeration variable of population density enters with a significantly positive coefficient, too. The number of college students, as a proxy of labour market quality and research activity in general is also positively related to firm location. One simple possible measure for agglomeration at the customer level, is the population density. Its sign is not straightforward. On the one hand, a more dense area allows for lower

[18]For example, in Figueiredo et al. (2002*b*) local wage has the expected sign, while in Holl (2004), the wage coefficient is insignificant. There are various explanations. For example, Figueiredo et al. (2002*a*) argue that firms consider the wage level as a determinant to locate in a cheaper country like Portugal (or even more so, Hungary) but within the country, wage has no effect.

transportation costs within the county, but on the other, it may lead to lower land prices and hence lower the cost of the investment. Results of the conditional logit are rather ambiguous (but other models suggest a positive relationship).

For these variables, cross correlations are interesting to look at. Distance to export destinations is negatively correlated with most development related variables. For over 80% of exports goes to Western Europe, this confirms a strong East-West division in Hungary. It is also clear, that Budapest is special in terms of these variables.

Model selection is not straightforward as explanatory power of the conditional logit models are fairly close. The McFadden's pseudo R^2 ranges between 19% and 20% when geographical dummies are inserted. Our preferred model is the one with regional and Budapest dummies, industry-specific wages and some control variables such as infrastructure ($CL4$).Likelihood-ratio tests confirm that control variables (LR chi2: 20, p=0.00) as well as dummies (LR Chi^2: 281, p=0.00) provide some significant additional information. (Given the geographic concentration of income, income variable is non-significant as the effect is picked up by dummies.) $CL4$ is preferred to $CL3$ owing to a better specified labor market.

1.4.3 Non-independent errors and the nested logit

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

$$p_j(y_j)/p_h(y_h) = \exp((y_j - y_h)\beta) \tag{1.4.5}$$

so that "relative probabilities for any two alternatives depend only on the attributes of those two alternatives" ((Wooldridge 2002*b*, 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.

To check whether the IIA assumption is strong enough, Hausman tests were run (Hausman & McFadden (1984)) for seven NUTS2 regions. Results (reported in Table 10.) show that the IIA assumptions almost always fail at the 1% level, suggesting that a more complex structure should be used. As is frequent for such exercise, asymptotic assumptions of the Hausman test fail for some occasions and hence, the generalized Hausman test was applied. Given that there is no theoretical support for having seven regions, so an alternative structure with three larger regions (West, Central, East) was drawn, and the tests were run only to indicate that IIA fails universally for such tree-structure.

One possible way to control for violations of the IIA assumption is to introduce dummy variables for each individual choice as suggested by Train (1986). Indeed, several specifications were run with fixed effects. To see if the introduction of fixed effects solved the problem, Hausman tests were re-run for a fixed effect specification. It did not solve the problem, all conditional logit structures may be rejected for violating IIA assumptions. This situation, often appearing in exercises similar to our own, requires the nested logit model to be called upon.

The nested logit model uses the same profit function as the conditional logit (1.4.1) but works with a decision tree. The firm now first picks a region out of upper level alternatives u, and then chooses a county within the already selected region, out of lower level alternatives, r. Importantly, no assumption on a two-step decision-making is necessary. It is enough to believe that certain counties are competing more closely than others.

Location probability in a county r, depends on probability of location in a region (u, upper level alternatives) times the probability of location in a county (m, lower level alternatives) in the given region:

$$\Pr_{ur} = \Pr_{r|u} * \Pr_{u}$$

$$\overset{NL}{\Pr_{r|u}} = \exp(\beta' Z_{ur}) / \sum_{n \in u} \exp(\beta' Z_{un})$$

where Z explains the choice of an upper level (region) alternative in the conditional logit case $\beta' Z = \gamma' b_r + \lambda' d_r^j$.

$$\Pr u = \exp(\alpha' W_u + \xi_u IV_u) / \sum_m \exp(\alpha' W_m + \xi_m IV_m) \tag{1.4.6}$$

In this last equation, the inclusive value, $IV = \ln(\sum_{n \in u} \exp(\beta' Z_{un}))$, will tell us if the nest helps. From Maddala (1983), we know that $0 \leq \xi \leq 1$ and when $\forall\ \xi_m = 1$, the NL collapses to CL, while if $\forall\ \xi_m = 0$, the upper nest matters only, i.e. firms choose a county randomly within the selected region.

It is important to stick to the RUM framework here as well, so a random utility maximization consistent nested logit had to be applied (Heiss (2002)). As a result, deterministic utilities must be scaled by the inverse of the IV_m parameters (ξ_m) in the conditional utility. This implies different scaling of the utilities across nests but allows the interpretation of $\beta' Z$ as RUM model.

$$\overset{RUMNL}{\underset{r|u}{\Pr}} = \exp(\beta' Z_{ur}/\xi_m) / \sum_{n \in u} \exp(\beta' Z_{un})$$

1.4.4 Nested logit results

There may be several natural nests: the seven NUTS2 regions, the three broad geographical areas: East, Central and West or our preferred 4 regions of East, West, South and the capital plus its neighbourhood. Results, reported in Table 8, provide solid support for many of our previous results. According to specification NL1, the basic variables: per capita income, size, local and national corporate access, business services access and wages, all enter significantly and with the expected sign. With disaggregated variables (specifications NL2-NL5), own industry output remains one of the best performing variables along with national (external) market access. Better local (internal) supplier access remains a point of attraction, too. National (external) access to suppliers and the own industry remain to enter with the negative sign. Other explanatory variables loose or gain significance depending on the nest.

Specification test of the nested logit model is based on the values of the inclusive value parameters. The LR test of homoskedasticity (all values equal one) is clearly rejected for all specifications. No single IV_m is ever close to the unity, suggesting that all parts of the nest is well warranted. However, greater than unity figures in general indicate some specification problem of the random utility framework. We checked for several possible nests - Table 8 reports results for three such nests - but failed to get inclusive values at or below unity. Once again, the model is likely to be misspecified, although (Train 2003, Chapter 4.) discusses studies that prove that for several cases, RUM may well be consistent with IV values above one.

Given the lack of a fully supported nest structure either by theory or by econometric evidence, model selection is difficult. We note that the addition of detailed access variables to specification $NL1$ makes a contribution, while no number of nests is clearly suggested by the data.

1.4.5 Count data methods and results

A great advantage of CL approach is its direct link with random utility maximisation. However, there may be several specification problems with the conditional logit model. The IIA assumption fails and the choice of a certain nested logit specification may seem somewhat arbitrary. Thus, one can apply count data models to see robustness of results. This comes with an additional advantage: the easy inclusion of time dummies. Indeed, during transition, there may have been important changes over time - such as shifts in public policy - affecting regions differently.

In an effort to check robustness of CLM, count data models are used in this section, with the dependent variable representing the number or frequency of a particular event, in our case, the number of investments in a particular county. In these models, coefficients explain why $x\%$ more projects took place in county A relative to county B.

Define $n^j_{r(t)}$ as the number of FDI investments in industry j, region r and time t. The explanatory variables are exactly the same as used in the previous sections.

$$Pr(Y^j_{r(t)} = n) = exp(-\lambda)\lambda n/n!$$

Importantly, Figueiredo et al. (2004) shows that the conditional logit equation as well as the Poisson model may stem from the same random utility maximisation model when firm-level characteristics are treated in a discrete fashion (such as operation in an industry). Alternative to the CL model, we can assume that $n^j_{r(t)}$ is the explained variable. These are independently Poisson distributed with

$$n^j_{r(t)} = \lambda^j_{r(t)} = \exp(\sum_j a^j d^j + \gamma' b_{r(t)} + \lambda' d^j_{r(t)})$$

where d^j are dummy variables indicating if a firm is in industry j. [19]

For every year, firm entry data were aggregated by industry and county, and Poisson regressions were run with the same set of explanatory variables used at logistic regressions (see Table 11). As expected, results were generally - but not always confirmed. Own industry output, once again proved to be one of the best performing variables with a coefficient close to 0.2, along with national (external) market access. However, supplier access variables swapped signs compared to logistic regressions. Other explanatory variables, such as distance from borders performed well, with even the number of college students making a difference. In a Poisson model context, the road network was unimportant while population density entered with a significantly positive variable suggesting the presence of some urbanization economies.

The Poisson model has the advantage of being closely related to the conditional logit, but it assumes that the conditional variance of the dependent variable, λ equals the conditional mean of λ. However, equidispersion is rare property in reality, and for most cases, the variance is larger than the mean. Overdispersion may be treated, but in a more general, negative binomial model that allows to test the null hypothesis of equidispersion.[20] Given their easy applicability, no wonder that both the Poisson and the negative binomial model have been used in location research (e.g. Basile (2004)).

The negative binomial distribution may be considered as a generalized Poisson, where the mean does not equal the variance. This deviation is represented with a dispersion parameter, α. The case with $\alpha = 0$ corresponds to equidispersion, and in that case the model collapses into a Poisson model.

Specification tests (LR test with one sided χ^2 statistics) suggested that the Poisson model is misspecified. However, results, reported as specification CNT5 and CNT6, suggest significance. In many cases even the magnitude of coefficients for the negative binomial are identical to those of the Poisson model despite the failure of the LR test. This robustness is not unusual in the literature, for example Smith & Florida (1994) finds a similar pattern for Poisson, negative binomial and even for the tobit model.

[19] Moreover, (Figueiredo et al. 2004, p. 203.) shows that the Poisson concentrated log likelihood is "identical to the conditional logit likelihood with some constraints."

[20] Importantly, the negative binomial model yields more efficient test statistics and prevents us from drawing overly optimistic conclusions (see Cameron & Trivedi (1998)).

Thus, we argue that results from the Poisson model, which may be derived from theory, may be considered adequate despite the lack of equidispersion.

1.4.6 Comparisons through time, entry mode and industry sector

This section compares results by three categories: mode of entry (greenfield versus foreign acquisition), time (early versus later years) and industrial sector (light industry versus equipment production). Table 12 provides details.

So far, we have looked at locational determinants of new firms only. We have data on foreign acquisitions that may either be considered as privatisation deals or investment by a foreign firm in an existing Hungarian company. Given that a substantial effort has been invested into linking firms that changed legal status but remained the same company in essence, our acquisition data include episodes when a firm exited and a new firm appeared at the same area with different ownership but very similar structure. There are altogether 870 foreign acquisitions, a quarter of which was certainly related to privatisation, i.e. a foreign share replacing state or municipal ownership.

Our regional fixed effect conditional logit regressions were run for both groups: new firms and foreign acquisitions. It turns out that several coefficients are broadly left unchanged - both in terms of sign and significance for these two groups. For example, the local presence of own industry remained a key determinant of location choice while some external access variables remained a strong deterrent. Some access variables, such as local market access, lost significance suggesting that for takeovers, the geography of own industry was the most important. Other explanatory variables such as wages or distance work equally well for the two groups; moreover, low wages seem to have been more important for companies purchased by foreign firms.

Pooled data for years have been used to access validity of our variables. It is interesting to see to what extent coefficients change through periods in time and across groups of industries. To see how variables evolved through time, fixed effect conditional logit regressions were run for two periods: 1993-1996 and 1997-2002. The first four years may be considered years of transition - that is when prices were fully liberalised, mass privatisation ended and foreign presence become a stable feature of the economy. Many coefficients, including those related to the input-output linkages changed little through time. However, high wages seem to have been more of a deterrent in the early nineties with the coefficient loosing significance after 2000 or skill-content, undetected by industry specific wages, has become more important of late.[21]

Given that the capital city Budapest is responsible for producing 35% of the GDP, its stance may be different for various modes of entry or time periods. Indeed, we find that new firms did actually have a preference for the capital, while this cannot be said about privatisation and acquisitions. One possible explanation is that newcomers had knowledge mostly about Budapest and this mattered when picking a field for a plant but had no role in the case of taking over an existing firm. The same information advantage may be appreciated by looking at the early years of the sample (transition, basically) versus the later years. Budapest mattered a great deal during the first few years, and its advantage halved for the second half of period.

Sectoral patterns may also be of interest both in terms of robustness of general results and because of a potentially mixed results for various sectors. Robustness of results was checked by running regressions and leaving out one industry at a time. Results varied marginally only. Second, some industries were grouped into two categories: light industry (e.g. textile, clothing, etc.) and miscallenous machinery (inc. electric machinery, audio-video equipment, vehicles, etc.), and regressions were run for one group at a time.

As was the case before, some key access variables remained unchanged. As for differences, light industries were more deterred by competition from regions in proximity: the national own industry output variable turned to be negative for the light industry group only, suggesting that nationwide competition was stronger for lower value added and/or less differentiated good producing sectors. At the same time, equipment producers - with a greater dependence on export markets - found distance from export destinations to be more important. Wages did matter a lot for the light industries, whereas higher skill-content sectors appreciated skills more.

[21] As a caveat here, note that comparison within a logistic framework is not directly possible. In a logit regression, the variance of the error term cannot be estimated together with parameters and as thus, the variance term is normalized to one. As a result, a difference in values may only be due to a difference in the variance of the error term. Hence a difference in the coefficient value may be distorted.

1.5 Conclusions and future research

This paper focused on location decisions of foreign investors within one country, using econometric models with discrete dependent variables that are generated from a tax report based dataset of Hungarian firms. The rapid appearance of foreign-owned manufacturing sites offered a great opportunity: studying the geographic properties of a large number of new firms entering a region previously closed to foreigners. Some conclusions may be drawn regarding theory and its empirical support as well as the validity of some methodologies.

Taking a snapshot of the economy rather than modelling long run equilibrium, one of our aims has been to bring a widely used class of new economic geography models to the data and investigate how well various channels of agglomeration and dispersion forces work. In the paper a possible way was shown to link input-output linkage based NEG theory and a tax report based dataset - building on variables that had been generated out of firm-level sales figures. In order to see validity of results, specifications of conditional logit, nested logit, Poisson and negative binomial models were tested. Although specification tests suggested that econometric models have generally been misspecified in one way or another, most coefficients kept their respective sign throughout specifications, and similar log likelihoods (or McFadden's pseudo R^2 measures, where available) suggested that most specifications are by and large equally supported by data.

Results that proved to be robust through discrete choice and count data specifications 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. The importance of industrial clustering has been robustly shown and some support of agglomeration externalities was found as well. Access to firms operating in the same industry as the new firm as well as proximity of potential customers throughout the country seemed to be a persistently important determinant of location choice. This provides some empirical support to NEG models with input and output linkages.

Wages have been important in explaining firm location. However, unless industry specific wages are used, the impact of labour costs is mostly undetected. Further, the addition of blue-collar wage costs that reflect the heterogeneity (in skills and training) of a relatively immobile and seemingly homogenous workforce improves our understanding.

The export distance measures are overwhelmingly significant with the expected negative sign in any specification. For a small and open economy this is not surprising. Most governments emphasise the construction of major East-West or North-South corridors and the importance of this notion is confirmed by the strong significance of our road distance to borders parameter. However, positive coefficient of the road network variable suggests that building roads within a county will foster FDI inflow as well.

Some important difficulties have arisen. First, the fact that a large share of action is going on within the own industry suggests that disentangling various agglomeration forces within an industry has once again proved to be rather difficult. As a result, when data permits, one would probably need to increase "data resolution" and leave two-digit industries (such as electronic equipment production) for three-digit sectors (e.g. medical equipment). Second, the unexpected sign of some access variables suggests that some important location force is missing (and is picked up some access variables) or the construction of access variables itself shall be reconsidered.

Finally, some broader policy conclusion may be drawn - with caution. 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. Further, subsidies to large firms may be efficient as long as they lure in similar firms. Second, input-output linkages are important. Thus, improving the relationship between suppliers and multinationals is key to fostering more investment. With a recent experience of loosing multinationals to non-EU Eastern Europe and China, this may be ever more important. Third, other explanatory variables that were found to be significant are telephone and road network, confirming the widely held view on the importance of local infrastructure.[22]

As for local policy, there is a narrow result, too. We have seen that a 10% increase in industrial output per county increases likelihood of the next firm location by 2.2% in the same manufacturing sector. Thus, FDI subsidies can be thought of a generator of future tax revenues by other firms. A subsidy to a key electronics firms that will raise output by 10% will come back as 2.2 average new firm will enter out of the potential 100. However, tax consequences are likely to be minor in absolute terms.

[22]However, one must bear in mind that several general equilibrium NEG models would show how construction of motorways may have an adverse effect in the long run. See Baldwin et al. (2003) for theory and Puga (2002) for some empirical support.

The fact that further suppliers are likely to be lured while others crowded out, the taxation consequence is even harder to quantify.

Table 1 New foreign manufacturing firms per Hungarian counties (1993-2002)

Counties	Inhabitants ('000)	Number of new firms*	New firms per capita ('000)	New firms during first five years	New firms during second five years
Szabolcs-Szatmár-Bereg	583	102	0.17	48	54
Borsod-Abaúj-Zemplén	738	139	0.19	73	66
Békés	393	79	0.20	45	34
Hajdú-Bihar	550	118	0.21	62	56
Jász-Nagykun-Szolnok	413	98	0.24	62	36
Somogy	334	104	0.31	73	31
Tolna	247	87	0.35	50	37
Nógrád	218	80	0.37	43	37
Heves	324	124	0.38	72	52
Fejér	428	168	0.39	111	57
Csongrád	426	184	0.43	107	77
Bács-Kiskun	542	236	0.44	146	90
Pest	1123	504	0.45	278	226
Veszprém	368	183	0.50	112	71
Zala	297	164	0.55	95	69
Baranya	402	235	0.58	147	88
Komárom-Esztergom	316	197	0.62	123	74
Vas	267	189	0.71	107	82
Győr-Moson-Sopron	440	346	0.79	213	133
Budapest city	1708	2013	1.18	1167	846

Source: KSH, APEH Corporate dataset, author's calculations

Table 2. Share of new firms by county categories

year	Budapest	richest 3	poorest 3
1993	39%	13%	5%
1994	44%	13%	4%
1995	37%	15%	5%
1996	32%	14%	6%
1997	32%	14%	6%
1998	37%	13%	8%
1999	39%	14%	8%
2000	39%	11%	7%
2001	37%	12%	6%
2002	38%	11%	6%

Source: APEH Corporate dataset, author's calculations

Table 3 New foreign manufacturing firms by industries

(NACE code) Industries	All FDI	Greenfield
(17) Textile	327	280
(18 & 19) Cloths, leather	452	397
(20 & 21) Paper and wood products	475	414
(22) Press	648	510
(23 & 24) Refinery and chemicals	208	156
(25) Plastic rubber	383	319
(26) Other non-metalic	283	229
(27) Metal -basic	68	54
(28 Metal -fabricated	725	602
(29 Machinery	632	525
(30 Office equipment	57	51
(31 Electric machines	208	179
(32 & 33) Audio-video, PC, etc. instruments	429	347
(34 & 35) Motor vehicles	153	133
(36) Furniture, etc.	302	261
Total manufacturing (ex-food)	5350	4457

Source: APEH Corporate dataset, author's calculations.

Table 4 Average unit transportation costs by industry

(NACE code) Industries	Unit price *
(17) Textile	11,6
(18 & 19) Cloths, leather	31,5
(20 & 21) Paper and wood products	5,8
(22) Press	22
(23 & 24) Refinery and chemicals	18
(25) Plastic rubber	12
(26) Other non-metalic	8
(27) Metal -basic	6
(28 Metal -fabricated	31
(29 Machinery	27
(30 Office equipment	140
(31 Electric machines	45
(32 & 33) Audio-video, PC, etc. instruments	140
(34 & 35) Motor vehicles	41
(36) Furniture, etc.	10
Total manufacturing (ex-food)	-

*Source: World Bank, APEH Corporate dataset, author's calculations. *Unit price in USD/kg - original World Bank data in ISIC terms, unit prices were transformed to NACE categories and aggregated by the author*

Table 5. Summary statistics

Variable	Description	Source	Mean	Std. Dev.
IPC	income per capita (Ft, '000)	KSH	87.2	27.6
Pop	population size ('000)	KSH	505	339
IPloc	own industry local output	APEH, „AKM" of KSH	649437	2637335
Ipnat	own industry national access	APEH, „AKM" of KSH	231341	501525
SAloc	local supplier access	APEH, „AKM" of KSH	1050664	2810798
MAloc	local market access	APEH, „AKM" of KSH	1879780	5910528
SAnat	national supplier access	APEH, „AKM" of KSH	354441	574277
MAnat	national market access	APEH, „AKM" of KSH	621667	1041976
BAloc	local business access	APEH, „AKM" of KSH	7269205	3058622
Tel_size	Size of telephone network (fixed line subscribers)	KSH	123244	158637
Road_size	Size of highway network (km)	KSH	1526.8	563.77
Edu_size	number of college students	KSH	9803	7332.54
Density	population density: inhabitants/area	KSH	0.24	0.69
dSouth	Distance of Southern export border (km)	HAS-Institute of Economics	254	117
dWest	Distance of Western export border (km)	HAS-Institute of Economics	233	100
dAirport	Distance of Airport (km)	HAS-Institute of Economics	136	684
Wage	local wage (Ft)	Minsitry of Labor „LMS"	31204	14371
Wage_ind	local, own industry wage (Ft)	Minsitry of Labor „LMS"	30362	16232
Wage_bc	local blue-collar wage (Ft)	Minsitry of Labor „LMS"	25585	12834
Wage_off	local office wage (Ft)	Minsitry of Labor „LMS"	38096	22391
Wage_man	local manager wage (Ft)	Minsitry of Labor „LMS"	80120	66476
D_{lr}	Road distance between cities (km)	HAS-Institute of Economics	190	103

KSH: Hungarian Central Statistics Office, „AKM": Input-output tables, „LMS": Annual Labour Market Survey, APEH: Hungarian Tax Authority's corporate database. NB All variables in estimations are taken in logs.

Table 6. Philips in Hungary

#	Sector	Address		Additional locations		
		zip-code	county	number of	zip-code	county
1	IT	8000	Fejér	1	1119	Budapest
2	Plastic	8660	Fejér	1	9600	Vas
3	Distribution	1051	Budapest	0		
4	Electronics	9700	Vas	1	9700	Vas
5	Audio	8000	Fejér	1	8000	Fehér
6	Electric machines	8000	Fejér	1	9027	Gyor-Sopron
7	Commerce, retail	1092	Budapest	0		
8	Commerce, retail	1119	Budapest	0		
9	Electric equipment	9700	Vas	0		

N.B. Status at 2002, excluding closed down entities. Source: Corporate Registry, author's collection.

Table 7. Conditional logit estimates

Specification	**CL1**	**CL2**	**CL3**	**CL4**	**CL5**	**CL6**
Fixed effects	no	no	1	1	20	20
Region FE	no	no	7	7	no	no
Ln (income per capita)	0.91***	0.03	-0.44	-0.16	-0.35	0.06
	(0.17)	(0.22)	(0.27)	(0.21)	(0.38)	(0.35)
Ln (population size)	0.20**	-0.19	-0.50	-0.43	0.01	1.47***
	(0.09)	(0.34)	(0.36)	(0.36)	(1.22)	(0.24)
Ln(own industry local output)	0.22***	0.20***	0.21***	0.21***	0.21***	0.21***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Ln (own industry national access	-0.08	-0.19***	-0.16**	-0.17**	-0.24***	-0.26***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Ln(local supplier access)	0.10**	0.03	0.11**	0.10*	0.11**	0.09
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
Ln(local market access)	-0.17***	-0.17***	-0.10**	-0.12**	-0.08	-0.09*
	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)
Ln (national supplier access	-0.35***	-0.88***	-0.62***	-0.60***	-0.99***	-1.04***
	(0.12)	(0.15)	(0.15)	(0.15)	(0.17)	(0.17)
Ln (national market access)	0.64***	0.41***	0.65***	0.60***	0.47***	0.40***
	(0.01)	(0.12)	(0.13)	(0.13)	(0.14)	(0.14)
Ln (local business access)	0.33***	-0.006	-0.10	-0.07	-0.17*	-0.14
	(0.05)	(0.07)	(0.08)	(0.08)	(0.10)	(0.10)
Ln (local wage)	-0.82**	-0.07	0.60		1.23**	
	(0.39)	(0.45)	(0.52)		(0.63)	
Ln (local, own industry wage)				-0.40**		-0.37*
				(0.19)		(0.19)
Ln (local blue-collar wage)				0.42**		0.37*
				(0.20)		(0.20)
Ln (number of college students)		0.51	0.78**	0.79**		
		(0.28)	(0.32)	(0.32)		
Ln (size of highway network)		0.25**	0.89*	0.98*	-0.83	-0.61
		(0.11)	(0.51)	(0.51)	(0.59)	(0.69)
Ln (size of telephone network)		0.18**	0.25***	0.19**		-0.10
		(0.09)	(0.10)	(0.09)		(0.13)
Ln (population density)		0.09				-0.42
		(0.07)				(0.61)
Ln (avg. distance export borders)		-0.62***	-0.53***	-0.60***	-1.40*	-0.67
		(0.13)	(0.14)	(0.14)	(0.84)	(0.44)
Ln (distance to Airport)		-0.62***				
		(0.13)				
Ln (distance to Western border)		-0.39***				
		(0.06)				
Ln (distance to Southern border)		-0.22***				
		(0.05)				
LR chi squared	4557	4557	4557	4310	4557	4310
Log likelihood	5187	5331	5347	5216	5440	5308
McFadden's pseudo R squared	0.19	0.19	0.20	0.20	0.20	0.20
Number of observations	-10759	-10686	-10678	-10433	-10631	-10387

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. Nested logit estimates

Specification	**NL1**	**NL2**	**NL3**	**NL4**	**NL5**	**NL6**
Top level alternatives	3	3	4	4	4	7
FE	NO	YES	NO	YES	YES	NO
Ln (income per capita)	0.67***	0.13	-0.19	-0.58	0.45	-0.30
	(0.18)	(0.55)	(0.31)	(0.52)	(0.38)	(0.35)
Ln (population size)	0.32***	2.67	1.74***	2.35***	2.61***	-0.53
	(0.13)	(1.90)	(0.42)	(0.59)	(0.63)	(0.58)
Ln (local corporate access)	0.18***					
	(0.04)					
Ln (national corporate access)	0.24***					
	(0.04)					
Ln(own industry local output)		0.37***	0.36***	0.35***	0.36***	0.28***
		(0.04)	(0.03)	(0.03)	(0.03)	(0.02)
Ln (own industry national access		-0.39***	-0.30***	-0.43***	-0.48***	-0.17**
		(0.11)	(0.10)	(0.11)	(0.11)	(0.07)
Ln(local supplier access)		0.21**	0.08	0.23***	0.19**	0.09*
		(0.09)	(0.08)	(0.08)	(0.08)	(0.05)
Ln(local market access)		-0.14	-0.18**	-0.04	-0.03	-0.05
		(0.09)	(0.07)	(0.08)	(0.08)	(0.04)
Ln (national supplier access		-1.51***	-0.96***	-1.45***	-1.58***	-0.74***
		(0.27)	(0.22)	(0.26)	(0.26)	(0.16)
Ln (national market access)		0.46**	0.84***	0.66***	0.60***	1.03***
		(0.22)	(0.19)	(0.22)	(0.22)	(0.13)
Ln (local business access)	0.29***	(-0.27	0.48***	-0.28	-0.35*	-0.02
		(0.19)	(0.12)	(0.18)	(0.18)	(0.11)
Ln (local wage)	-1.34***			1.88**		0.91
	(0.45)			(0.84)		(0.69)
Ln (local, own industry wage)		-0.32	-0.42**		-0.38**	
		(0.22)	(0.21)		(0.22)	
Ln (avg. distance export borders)		0.80	-1.52***	0.19	0.96	
		(1.75)	(0.33)	(0.88)	(1.08)	
Ln (size of highway network)		-0.29				0.23*
		(0.96)				(0.11)
Ln (number of college students)						1.38***
						(0.52)
Ln (Size of telephone network)		-0.11				0.21*
		(0.22)				(0.11)
Inclusive value 1	0.96	1.78***	1.99***	1.96***	1.90***	1.46***
Inclusive value 2	1.10*	3.33***	3.09***	2.87***	4.41	1.84***
Inclusive value 3	0.86**	1.95***	2.99***	3.37**	1.89***	2.28***
Inclusive value 4			2.82***	1.87***	3.53***	2.11***
Inclusive value 5						1.51***
Inclusive value 6						1.16
Inclusive value 7						1.71***
Method	NLRUM	NLRUM	NLRUM	NLRUM	NLRUM	NLRUM
Number of observations	4457	4412	4412	4457	4412	4457
Model LR chi2	4964	5379	5235	5490	5385	6505
Log likelihood	-10869	-10527	-10599	-10606	-10524	-12774
LR test of IVs=1	76.05	49.18	95.71	51.68	56.9	103.8
	0.00	0.00	0.00	0.00	0.00	(0.00)

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9. Cross correlation of variables

Access variables

	Ln (own ind loc acc)	Ln (own ind nat acc)	Ln (local supp acc)	Ln (local market acc)	Ln (nat supp acc)	Ln (nat market acc)
Ln (own industry local output)	1.00					
Ln (own industry national access	0.57	1.00				
Ln (local supplier access)	0.53	0.66	1.00			
Ln (local market access)	0.60	0.52	0.33	1.00		
Ln (national supplier access	0.42	0.74	0.32	0.70	1.00	
Ln (national market access)	0.41	0.48	0.58	0.41	0.55	1.00

Development related variables (including Budapest)

	ln (income per capita)	Ln (bus. service acc)	LN (local wage)	Ln (distance border)	Ln (telephone netw	Ln (highway netw)
ln (income per capita)	1.00					
Ln (business service access)	0.54	1.00				
LN (local wage)	0.11	0.71	1.00			
Ln (weighted distance of export borders)	-0.56	-0.24	-0.11	1.00		
Ln (size of telephone network)	0.43	0.93	0.67	-0.14	1.00	
Ln (size of highway network)	-0.20	-0.30	-0.02	-0.37	-0.34	1.00

Development related variables (excluding Budapest)

	ln (income per capita)	Ln (bus. service acc)	LN (local wage)	Ln (distance border)	Ln (telephone netw	Ln (highway netw)
ln (income per capita)	1.00					
Ln (business service access)	0.09	1.00				
LN (local wage)	-0.02	0.84	1.00			
Ln (weighted distance of export borders)	-0.63	-0.18	-0.08	1.00		
Ln (size of telephone network)	-0.03	0.90	0.75	-0.04	1.00	
Ln (size of highway network)	0.21	0.05	0.06	-0.52	-0.05	1.00

Table 10 Generalised Hausman tests of IIA

	$\chi 2$ test (p-value)	
7 NUTS2 regions	No county fixed effects	With county fixed effects
All versus no Region1	147.34***	209.79***
	(0.00)	(0.00)
All versus no Region2	76.03***	47.59***
	(0.00)	(0.00)
All versus no Region3	41.18***	69.68***
	(0.00)	(0.00)
All versus no Region4	39.83***	44.28**
	(0.00)	(0.01)
All versus no Region5	19.05	25.80
	(0.161)	(0.47)
All versus no Region6	49.39***	51.68***
	(0.00)	(0.00)
All versus no Region7	37.86***	34.83
	(0.00)	(0.11)

	$\chi 2$ test (p-value)	
3 large regions: West, East, Central	No county fixed effects	With county fixed effects
All versus no West	96.34	143.98
	(0.00)	(0.00)
All versus no Central	60.80	79.89
	(0.00)	(0.00)
All versus no East	106.36	43.65
	(0.00)	(0.00)

Table 11. Location choice with count data regressions

Specification	CNT(1)	CNT(2)	CNT(3)	CNT(4)	CNT(5)	CNT(6)
Model	Poisson	Poisson	Poisson	Poisson	NegBin	NegBin
FE	No	No	County	Area, time	No	Area, time
Ln (income per capita)	1.62***	0.65***	-0.10	0.39***	0.92***	0.53***
	(0.11)	(0.15)	(0.26)	(0.12)	(0.19)	(0.16)
Ln (population size)	0.82***					
	(0.07)					
Ln(own industry local output)	0.23***	0.24***	0.25***	0.25***	0.26***	0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ln (own industry national access)	-0.02*	-0.03**	-0.04***	-0.03**	-0.12***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Ln(local supplier access)	-0.09***	-0.12***	-0.07***	-0.14***	-0.17***	-0.18***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Ln(local market access)	-0.06***	-0.08***	0.02	-0.01	-0.05*	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Ln (national supplier access)	0.02	0.07***	0.04	0.12***	0.25***	0.29***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)
Ln (national market access)	0.17***	0.08***	0.01	0.13***	0.05*	0.08***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Ln (local business access)	0.006	-0.10**	-0.13***	0.47***	-0.12*	0.43***
	(0.04)	(0.04)	(0.03)	(0.02)	(0.06)	(0.03)
Ln (local wage)	-0.86***					
	(0.08)					
Ln (local, own industry wage)		-0.68***	-0.58***	-0.81***	-0.76***	-0.86***
		(0.06)	(0.06)	(0.07)	(0.08)	(0.09)
Ln (local blue-collar wage)						
Ln (number of college students)		0.73***			0.63***	
		(0.07)			(0.09)	
Ln (size of highway network)		0.05			0.02	
		(0.04)			(0.06)	
Ln (population density: inhabitants/area)		0.11***			0.13*	
		(0.05)			(0.06)	
Ln (size of telepohone network)		0.11*			0.18**	
		(0.06)			(0.09)	
Ln (avg distance export borders)		-0.48***			-0.40***	
		(0.05)			(0.07)	
Ln Distance of Airport						
Ln Distance of Western export border						
Ln Distance of Southern export border						
LR $\chi 2$	6936	6661	6830	6837	1748	1822
Log likelihood	-4912	-4579	-4494	-4491	-4163	-4125
McFadden's pseudo R2	0.4138	0.4210	0.4318	0.4322	0.1735	0.1809
Over-dispersion α +					0.36	0.33
LR (α=0), $\chi 01$ (p-value)					833 (0.00)	730(0.00)
Number of observations	3000	2737	2737	2737	2737	2737

Standard errors in parentheses. Significance at 1%, 5% and 10% is denoted by ***, **, and *, respectively
+ $\chi 01$: is a one-sided $\chi 2$ test of the over-dispersion parameter, α.

Table 12. Comparisons through mode of entry, time and sector

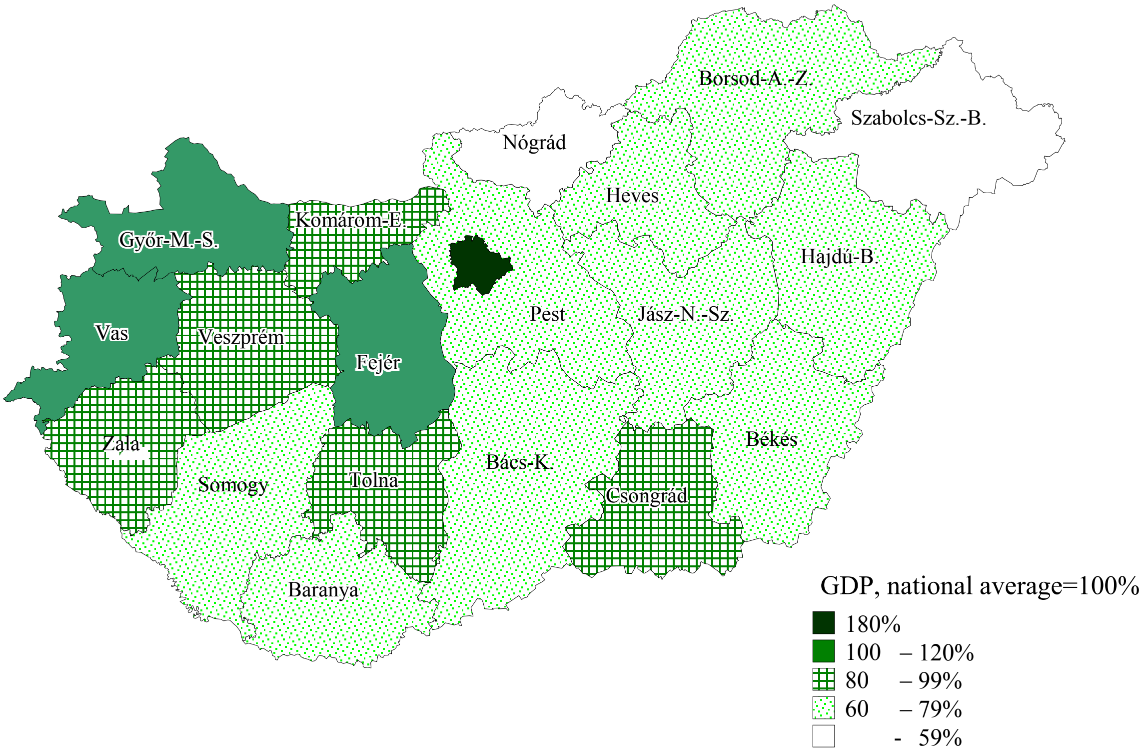
Specification	CLC1	CLC2	CLC3	CLC4	CLC5	CLC6
Subject of comparison	Mode of entry		Time period		Industry sector	
	foraq=0	foraq=1	1993-1997	1998-2002	light ind.	equipment
Ln (income per capita)	-0.02	0.28	-0.64	0.02	0.61*	0.53
	(0.22)	(0.53)	-0.51	-0.33	-0.36	-0.49
Ln (population size)	-0.32	-1.49*	-1.03*	-0.61	-0.23	-0.13
	(0.37)	(0.88)	-0.61	-0.55	-0.59	-0.81
Ln(own industry local output)	0.21***	0.21***	0.24***	0.20***	0.31***	0.15***
	(0.02)	(0.05)	-0.03	-0.03	-0.04	-0.04
Ln (own industry national access	-0.17**	-0.14	-0.16	-0.19**	-0.38***	0.17
	(0.07)	(0.15)	-0.1	-0.1	-0.14	-0.18
Ln(local market access)	-0.13***	0.12	-0.07	-0.13**	-0.13*	0.04
	(0.05)	(0.10)	-0.07	-0.07	-0.07	-0.12
Ln(local supplier access)	0.09*	-0.09	0.27***	0.05	-0.09	-0.15
	(0.05)	(0.12)	-0.09	-0.07	-0.09	-0.13
Ln (national market access	0.58***	0.81***	0.60***	0.59***	1.03***	0.98***
	(0.13)	(0.29)	-0.19	-0.18	-0.23	-0.36
Ln (national supplier access)	-0.63***	-1.46***	-0.1	-0.97***	-0.34	-0.89***
	(0.15)	(0.34)	-0.24	-0.21	-0.3	-0.33
Ln (local business access)	-0.01	-0.20	0.07	0.09	-0.02	-0.03
	(0.08)	(0.21)	-0.17	-0.11	-0.14	-0.18
Ln (local, own industry wage)	-0.40**	-1.03**	-0.77**	-0.42*	-1.01***	0.44
	(0.19)	(0.44)	-0.35	-0.23	-0.28	-0.46
Ln (local blue-collar wage)	0.42**	0.72	0.73**	0.50**	0.46	0.4
	(0.20)	(0.44)	-0.33	-0.25	-0.29	-0.52
Ln (avg. distance export borders)	-0.74***	-1.20***	-0.45*	-0.84***	-0.15	-0.65*
	(0.15)	(0.38)	-0.26	-0.22	-0.25	-0.38
Ln (number of college students)	0.66**	2.52***	0.21	1.18**	0.89*	0.38
	(0.33)	(0.80)	-0.51	-0.48	-0.51	-0.73
Ln (size of telephone network)	0.20**	-0.16	0.88***	-0.17	0.31**	0.33*
	(0.09)	(0.20)	-0.19	-0.26	-0.15	-0.2
Ln (size of highway network)	2.72***	-0.02	3.87**	2.45*	3.97***	1.81
	(0.93)	(2.09)	-1.62	-1.39	-1.41	-2.04
Budapest dummy	2.35***	-1.47	3.17***	1.90*	3.96***	1.68
	-0.71	-1.56	-1.09	-1.05	-1.08	-1.52
Fixed effects	Bud	Bud	Bud	Bud	Bud	Bud
Region FE	yes	yes	yes	yes	yes	yes
entry mode	greenfield	for. acq.	greenfield	greenfield	greenfield	greenfield
time period	1993-2002	1993-2002	1993-1997	1998-2002	1993-2002	1993-2002
sector	all	all	all	all	18,19,20,21	29,31,31,32
Observations	4310	870	2011	2300	1890	1040
Model chi-square	5221	1200	2612	2667	2716	1414
df	23	23	23	23	23	23
Log likelihood	-10431	-2031	-4778	-5623	-4334	-2433
Pseudo R2	0.20	0.23	0.21	0.19	0.24	0.23

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1

Relative GDP per capita in Hungarian counties in 1998



Chapter 2

Location of FDI in Hungary: input-output linkages and the role of local policies

Abstract
The paper tests the effect of local development, regional and local policies on the location decisions of foreign owned firms in Hungary. Development indicators include for example local research and education activity while policy decisions used in this study encompass tax rates, investment incentives or road construction. The study builds upon a large national panel of firms and a recent survey of municipal policies. Among others, we find that density of road network, certain local concessions and a somewhat larger size of administration did positively influence location choice.

2.1 Introduction

Various economic policy decisions influence corporate behaviour: measures may be taken at municipal, regional or national level. From a policy perspective, corporate reaction to economic policy measures and development incentives has a great significance. First, European Union development policies will either target firms directly or influence them as a by-product of cohesion efforts. Second, while several empirical papers discuss the linkages between firms and policy in the US or Western Europe, the topic has rarely been discussed in the context of a less developed country.

In particular, we are interested in the effect of local development as well as regional and local policies on the location decisions of firms. We argue that policies should, *inter alia,* be evaluated on the basis of their impact on improving the economic environment and business conditions for firms in manufacturing. State involvement in the economy is related to the provision of public goods (such as the road network) and promotion of economic activities. Measures would influence other actors of the economy, such as nearby firms or related industries.

Development indicators, which would capture these externalities, also appear as cost factors for firms, and include local research and development activity, telephone network or education. Policy decisions used in this study encompass for example tax rates, municipal and national investment expenditures or investment incentives.

The study builds upon a large national panel of firms. Rather than following sectoral patterns, this dataset allows to analyse firm behaviour directly. We use several sources for policy and development variables, and a recent survey on municipal policies is used in addition to previously available sources of county-level data.

The paper is organised as follows. First, we give a brief summary of the key theoretical underpinnings of the research as well as surveys of the empirical findings of previous papers. This is followed by a presentation of the econometric model along with a description of the datasets and variables in section three. Section four present the results and some conclusions are drawn in section five.

2.2 Related literature

In this paper, we focus on location choices of foreign firms in manufacturing. Such companies have chosen relatively small number of locations for their production between 1991 and 2003. Indeed, a key stylized fact is the presence of concentration at the national level: a few counties attracted the majority of investment.

In models of economic geography, location decisions would be based on three types of factors. Choices will be influenced by some geographic properties of counties such as size or presence of mountains and rivers - this is called the *"first nature" geography* and these features would change very slowly or not alter at all. Location choices as well as behaviour of other firms determine *"second nature" geography* with sales between various firms becoming a key pulling factor. The second nature properties of a county may change rather quickly - as the experience of transition in Central Europe would suggest. In addition to these, location choices are influenced by the "*one and a half nature*": roads, universities or administration capacities that change more slowly than firms' activity but nevertheless, adapt to corporate needs as well as shape firms' behaviour.[1] This paper tries to evaluate the impact of the this "one and a half nature" features - while emphasising the important role "second nature" geography plays.

The empirical investigation is built upon a group of new economic geography models using input-output linkages among firms[2]. We assume that location choice of firms at the county level is unaffected by first nature geography in a small and plain country like Hungary. Labor costs are of course a key determinant, and second nature geography is important: various market access and agglomeration variables will be created and used as regressor. One and a half nature geography will be investigated in detail including regional development, accessibility and transportation network as well as local taxes and policies. Before turning to the analysis of our data, let us survey empirical evidence from various developed and less developed countries and note some recent results from Hungary.

2.2.1 Agglomeration and market access

Most models of new economic geography (or NEG) aim at uncovering the essential reasons behind both agglomeration and dispersion of economic activity by taking into account "second nature" geography features, such as access or proximity to potential consumers as well as suppliers of intermediate goods necessary for production.

As for the access to markets, the key idea that firm location depends on the proximity of demand was introduced a long ago, and already in 1954, Harris devised the simplest aggregate market-potential function. Market potential has been first investigated in an international context; proximity to key markets and suppliers has been explicitly featured in empirical works explaining overall economic activity or per capita income. Redding & Venables (2004) argue that a country's wage level (proxied by per capita income) is dependent on its capacity to reach export markets and necessary intermediate goods cheaply.

Agglomeration externalities were first emphasised by Marshall, and formalisation

[1] For more on this, see Baldwin (2004)

[2] A detailed description of a Krugman & Venables (1995) type model may be found in Békés (2005). An excellent survey of key hypotheses emerging from models of new economic geography and their mixed empirical support can be found in Head & Mayer (2004).

of most such externalities may be found in (Fujita et al. 1999, Ch. 16.). One such agglomeration force is labour migration: an increased population generated greater demand inviting more firms to settle in a larger city, and this allowed for a lower import bill and hence, lower living costs in general. Another driver of co-location of firms comes from the potential of supplier-buyer link between firms, i.e. one firm's output is the intermediate good of another as in Krugman & Venables (1995). Thus, firms try to locate close to other firms, hence lowering transaction costs. An other reason for agglomeration is the presence of knowledge spillover: proximity allows to exchange inventions while technology spillovers help increase productivity using other firms' knowledge. Further, labour pooling may be important as firms would enjoy the presence of a larger set of labour pool where the specific knowledge required by the firm can be fished out easily (Amiti & Pissarides (2001)).

There have been several papers dealing with location decisions of foreign investors and clustering of these firms. Head & Mayer (2005) look at Japanese investments carried out in the European Union; results show that market potential measures as well as agglomeration variables turn out to be significant determinants. Crozet et al. (2004) study location of FDI in France and 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. Similarly, a study by Head & Ries (2001) looks at Japanese investments in the US and finds that firms belonging to the same *keiritsu* tend to settle close to each other. Some studies considered countries of similar size and population to Hungary, for example Barrios et al. (2003) look at multinationals' location choice in Ireland to find that agglomeration forces contributed substantially to location choices but proximity to major ports and airports was also helpful.

Urbanization, i.e. density of the actual location may foster agglomeration by helping face-to-face communication or the spillover knowledge. Of course, high land prices and congestion may be a deterrent factor. Coughlin & Segev (2000) found a positive effect of urbanisation on location of manufacturing plants. Proximity to businesses that provide services for manufacturing firms such as banks or accountancies has been shown to attract investments.

2.2.2 Development, accessibility and transportation network

In a broader sense, regional development has often been investigated. For example, Basile (2004) showed that public infrastructure and education are attracting forces while crime rate is negatively related to new investments in Italy. Several studies considered the role of transportation per se in a regional setting. Cieslik (2003) looked at 50 Polish regions to find that both proximity of main export targets and road network have been the key magnets for foreign investment.

In the lack of appropriate data, only a few studies investigated the role of settlement level determinants of location choice. Holl (2004) analysed explicitly the impact of road infrastructure on new manufacturing establishments in Spanish municipalities. The paper suggests that infrastructure development affects municipalities differently even within one region and agglomeration forces operate within a relatively small geographic scope. Holl posits that a new motorway will positively affect productivity of firms in the very proximity of the motorway but adds that a negative spillover to more distant areas is likely as they loose out on investments. Results suggest that apart from the size of the settlement, share of educated workforce and proximity to major cities attract new investments, while competition presents a deterrent force. In contrast with other studies showing positive spill-over effects from co-location within a region or country, at a lower level of aggregation, competition overweighs these externalities. Most interestingly, it is shown that there is an average 14% increase in firm entry for municipalities located within 10km from the new motorway. Outside this 10km corridor, distance from motorways plays a small role only. Woodward (1992) took local transportation linkages as a separate variable to measure accessibility of regional and national markets. Here, interstate highway connection was taken as proxy to good access, and the positive and significant coefficients confirmed hypotheses.

Another way to look at transportation infrastructure is to estimate the impact of road density. A more developed network should help firms trade with other companies in the neighbourhood as well as transport final goods to cities. Hence, good transportation within regions allows for agglomeration externalities to yield greater profits form specialisation and economies of scale or technological spill-overs. In Indonesia for example, Deichmann et al. (2005) found that road density positively influences location choice for most of the industries. For China, Amiti & Javorcik (2003) found strong evidence of the importance of railway network.

By theory, the impact of access to key transportation channels may not serve as an attraction force. Recent models of new economic geography[3] suggest that providing a new transportation link between a rich and a poor region may exacerbate agglomeration tendencies, leading to new investment in the agglomerated (richer) area and hence, a greater divergence.

2.2.3 Labour market

In previous studies, various labour market variables have been investigated including gross wages, income tax rates, unemployment or the composition and skills of the labour force. Theoretically, lower wages reduce production costs and higher unemployment provides the necessary labour supply for new investments, thus, both should attract FDI. Studies of international location choice certainly support this position, while results are quite mixed when considering intra-national choice. For example, in Figueiredo et al. (2002*b*) local wage has the expected sign, but in other studies like Holl (2004), the wage coefficient is insignificant.

There may be various explanations for ambiguous results. Labour migration within one country may be strong thus alleviating differences. Different industries would use different types of labour in terms of skills and profession. The share of blue-collar workers may vary a great deal among sectors and furthermore, their wage may differ greatly depending on how skilled they are. Hence, the industry profile of a region may well influence the average wages. 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.

2.2.4 Local taxes and regional policies

There have been a few studies looking at local and regional taxes as well as regional policy initiatives. Woodward (1992) analysed a period of booming Japanese FDI activities in the US, focusing on greenfield start-ups that, unlike foreign acquisitions, require an explicit location decision. Location of 540 plants were analysed with firms assumed to have freely chosen a US state and a county. Interesting explanatory

[3]See Baldwin et al. (2003), Head & Mayer (2004) or Martin (1999).

variables include various tax rates, the presence of industrial policy (at the state level) and manufacturing agglomeration, racial and educational mix of population and labour market features (at the county level). High taxes did serve as a deterrent at the state level but the local property tax seemed to have no direct effect.As for the county level regressions, labour market variables proved to be important determinants of location choice.

Measuring state policy towards FDI was not easy. Woodward (1992) used an index developed by Luger (1987) and it included land and building subsidies, debt and equity capital support, job training, infrastructure improvement and site preparation. Another instrument is the presence of state-level investment and export promotion offices operating in Japan. In the early eighties only 15 US states had such office, but by the end of the decade most states had established such institution. Interestingly state effort had no significant impact while, an office in Japan proved an efficient tool to attract investments. For the late eighties and early nineties, Kim et al. (2003) considered new manufacturing FDI plants in the US to analyze the effect of industry promotion programs[4] by states. The impact of expenditures on FDI attraction programs was estimated and found to have a positive and significant effect. Moreover, Kim et al. (2003) suggested that promotion expenditures may be used to offset the lack of agglomeration.

Another way of looking at regional policy is to consider national initiatives to attract FDI into certain areas of the country. Barrios et al. (2003) find evidence that higher public incentives in Irish designated areas have increased the probability of multinational investment. In the United Kingdom, Devereux et al. (2003) examine whether discretionary government grants influence firm location. It is found that policy instruments in the form of regional grants do have some effect in attracting new firms to supported location, but this effect is rather small.

In several Central and Eastern European countries, special industrial zones were created to attract foreign investors. Several studies argued that zones would have a favourable impact. However, for Poland, Cieslik (2003) found that when controlling for access and agglomeration variables, the existence of such zones had no considerable impact on the number of investments.

Spending on incentives and infrastructure should have a favourable impact, but bureaucracy as a potentially important impediment to investment must be taken into

[4]In the US, there exists a central database, the "State Export Program Database" that collects state programs prepared by the National Association of State Development Agencies.

account as well. Deichmann et al. (2005) investigated the impact of local bureaucratic costs of doing business in Indonesia and found that the occurence of local interventions has a small negative effect, especially for regulation sensitive industries, such as tobacco.

Although local taxes have been found to be a deterrent force for firms (Bartik (1985), Papke (1989)), sensitivity was often found to be rather low and highly variable among industries and firm size (Freidman et al. (1992)). Looking at growth of establishments in Maine (USA), Gabe (2003) found that the local (personal) property tax rate has a negative effect on establishment growth but local government expenditure variables show little or no correlation with firm development. Local taxes in particular have an adverse effect, but the coefficients are almost negligible in size.

2.2.5 Hungarian results

Agglomeration of investments and a spatial polarization have also been visible phenomena in many sectors. For example, manufacturing of electronic devices by firms in Central and Eastern Europe 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[5].

To our knowledge, the impact of such variables on firm location in Hungary has not been investigated in detail. However, various agglomeration forces have been described and shown to be in work in Hungary and several policy and infrastructure variables were used to explain development patterns.

Barta (2003) described regional differentiation in post-transition Hungary giving a good example of agglomeration forces in work in the automotive industry. 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.

There have been several studies discussing the role of accessibility in influencing municipal and regional development in Hungary. Németh (2004) examined which variables could explain income per levels and unemployment rates in NUTS4 “kistérség” regions. Unemployment rates were substantially lower in regions close to the Western

[5] For details see Barta (2003).

borders as early as 1990 and the East-West division remained an important explanatory variable throughout the nineties. Apart from the usual measures of income (education or age), proximity to the capital city as well as the Western border have been key in explaining higher wages. Proximity to other borders proved to be insignificant.

Fazekas (2003) is closer to this research as it considers FDI and not development in general. In the focus of the paper lays the impact of FDI from a labour market perspective to study the impact capital inflow had on the regional structure of the country. The paper finds that concentration pattern of foreign-owned enterprises is just marginally higher than that of the domestically owned ones. However, FEs are concentrated in a different pattern, being located closely to the Western border. The approach of this paper is somewhat different to Fazekas (2003) in that it uses firm level data and investigates the agglomeration patterns of foreign firms only.

2.3 Econometric model

2.3.1 The estimation problem

Firms choose a location by maximising the (expected) profit function that depends on several explanatory variables. For parsimonious notation, let us introduce $INC_{r(t-1)}$ as the measure of county level income, the vector of variables $ACC^{j}_{r(t-1)}$ that includes all industry specific access variables and $wage^{j}_{r(t-1)}$ for county level wages. Further, all county level aggregate measures (such as the size of road network or university students) are included in the $County_{r(t-1)}$ vector. Survey based averages of municipal level policy variables are captured in the $local\_avg_{r(t-1)}$ vector. (see details below). As a result, our expected profit function for a firm i is:

$$\begin{aligned} \pi^{j}_{r(t)}(i) = \alpha_1 wage^{j}_{r(t-1)} &+ \alpha_2 INC_{r(t-1)} + \beta^{'}_{\mathbf{1}}\mathbf{ACC}^{j}_{r(t-1)} + \\ &+ \gamma County_{r(t-1)} + \gamma local\_avg_{r(t-1)} + \zeta^{j}_{r(t)}(i) \end{aligned} \quad (2.3.1)$$

where the error term, $\zeta^{j}_{r(t)}(i)$ includes all the non-observed variables.

Note that explanatory variables that have a time dimension are lagged one year. 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 try to avoid endogeneity, and lagging will free the model of simultaneity bias. In addition to lagging, we also need to assume that firms at time t considering values of explanatory variables at time $t-1$, pick a county independently of each other. Agglomeration works as firms locate close to other firms that had settled previously, but there is *no* strategic interaction between firms settling at time t. This is a necessary assumption for using simple discrete choice model.

In our econometric structure, firms base their location decision on expected profits conditional on choosing a particular location, but they make errors due to unobserved features of the various regions/settlements as well as inability to make perfect decisions. However, the likelihood of choosing a particular location does indeed depend on the expected profit there. This gives the basis of the Random Utility Maximisation (RUM) models such as ours. The econometric model that follows from RUM models is the McFadden (1974) type conditional logit. However, for several setups, it may be shown to be equivalent to the Poisson model and (Figueiredo et al. 2004, p. 203.) shows that the Poisson concentrated log likelihood is "identical to the conditional logit likelihood with some constraints."[6] Given their easy applicability, no wonder that both the Poisson and the negative binomial model have been used in location research.[7]

In our count data models, the dependent variable represents the number or frequency of a particular event, in our case, the number of investments in a particular county for a given year and industry. In these models, coefficients explain why $x\%$ more projects took place in county A relative to county B.

Define $n^j_{r(t)}$ as the number of investments in industry j, region r and time t. The expected value of the number of projects is:

$$E(n^j_{r(t)}) = \lambda^j_{r(t)} = \exp(\beta\prime X^j_{r(t-1)}) \tag{2.3.2}$$

The probability of the actual number of investments being $n^j_{r(t)}$ is:

$$Pr(n^j_{r(t)}) = \frac{e^{(-\lambda^j_{r(t)})}(\lambda^j_{r(t)})^{n^j_{r(t)}}}{n^j_{r(t)}!} \tag{2.3.3}$$

[6] One advantage of count data models is their applicability for large choice sets. In this paper, we use count data models to get results that may later be comparable with results on settlement level decisions.

[7] For example, see Basile (2004), Holl (2004)

where the Xs are the explanatory variables. For every year, firm entry data were aggregated by industry and county, and Poisson regressions were run with the same set of explanatory variables used at logistic regressions.

The Poisson model has the advantage of being closely related to the conditional logit, but it assumes that the conditional variance of the dependent variable, λ equals the conditional mean of λ. However, equidispersion is a rare property of firm level data, and for most cases, the variance is larger than the mean. Overdispersion may be treated, but only in a more general, negative binomial model that allows to test the null hypothesis of equidispersion.[8] The negative binomial distribution may be considered as a generalized Poisson, where the mean does not equal the variance. This deviation is represented with a dispersion parameter, α. The case with $\alpha = 0$ corresponds to equidispersion, and in that case the model collapses into a Poisson model.

2.3.2 Data and variables

To study location choices, we distinguish four types of forces. First, classic variables include gravity type variables (size, income par capita) and labour market measures as well as economic geography variables that are centered around two key determinants of location: agglomeration externalities and market access. Second, we use several municipal and regional infrastructure and development variables. Third, policy variables from the municipality survey (such as local tax rates) are included. Due to data availability, empirical results in this paper are based on county level data. Table 1 summarizes variables for this county level exercise.

Basic determinants and access variables

Classic determinants include the measure of income per capita as well as labour market features such as the average regional wage. To measure consumer demand, two variables were created as the total income is taken as income per capita multiplied by size of population. In addition to this, foreign demand is estimated with a proxy of access to foreign markets. Wage is measured by the average county level wage or the industry specific county level wage.

[8]Importantly, the negative binomial model yields more efficient test statistics and prevents us from drawing overly optimistic conclusions (see Cameron & Trivedi (1998)).

Economic geography variables are based on the concept of market access that posits that firm location depends on the proximity of demand. Building on the previous chapter, input-output linkages between firms are taken into account and several corporate access variables are estimated, including access to suppliers and corporate customers. These access variables measures proximity to firms that may be relevant for a new company, and the access variable is sum of output by firms weighted by distance and share in inter-company trade.

From theory, we would need one variable to measure demand and another one to proxy supplier access. Bear in mind, that although supplier and market access variables are compiled in a similar fashion here, they measure different types of variables: the market access is about demand, while the supplier access is just a proxy to (intermediate goods) prices. Further, both variables are divided into two components: one to pick up access to local (internal or within county) firms and another one for non-local (external or outside the county) firms. The reason for such dichotomy comes from the suspicion that the effect of distance is not linear, and firms clustered in one city or in a few cities close to each other, enjoy special agglomeration effects.[9].

Theoretically these are the basic access variables we need. However, there may be (and there is indeed) a strong correlation between supplier and market access variables. One possible reason for correlation between access variables is the fact that own industry(i.e. the very manufacturing sector of the new firm that is to choose a location) output influences both the supplier and the market access variable strongly. This stems from the structure of commerce between firms: companies trade the most with other companies in the very same industry. This feature makes the use of models with two sectors, such as upstream and downstream industries, impossible. On average, intra-industry trade amounts to one third of total inter-company sales, and this exacerbates correlation between the relevant supplier and market access variables. To remedy this, two new variables are introduced to measure separately the industry output of the actual firm: *local_own_industry* and *national_own_industry*. Output by this own industry are excluded in all other variables applied: *local_suppliers* and *local_markets*, as well as *national_suppliers* and *national_markets*.

For a formal description of the access variables, see subsection 1.3.2 of the previous Chapter.

Access to foreign markets influencing both demand and intermediate good prices,

[9] In a somewhat similar setup, Amiti & Javorcik (2003) created such aggregate access variables.

is measured by a single variable: $foreign\_market\_distance$. This takes into account that export is a crucial determinant of the revenue of Hungarian firms and the average import share reached 34% for manufacturing. Due to data limitation problems, this paper proxies access to foreign markets by weighing distances to the key export borders. Finally, $business\_services$ picks up access to local business services, as a special determinant of production costs.

Note that agglomeration externalities such as technology spill-over are not measured directly, but proxied by the local output of the given industry. Some other NEG models incorporated input and/or output competition among firms (Baldwin et al. (2003)). In our case, input competition, which is expected to be important locally, is picked up by local access variables, while output competition, which shall be more important nationally, would be indicated by a significantly negative national own-industry access variable.

Municipal and regional infrastructure and development

We know that a better municipal infrastructure in general lowers transaction costs of firms operating at or in the proximity of the actual settlement. As a result, higher expected profits should attract more firms in the areas. Below, variables of the vector $County_{r(t-1)}$ are discussed.

First, we take Hungarian Statistics Office (KSH) data on county aggregates to measure human infrastructure such as the presence of research activity and administration capacities. R&D is measured by the number of research centers (at universities or elsewhere), the number of employees at such centers, and the annual expenditures at these centers. The role of universities is also captured by the number of students enrolled at high education institutions in the given county. All these variables are taken relative to the population of the county.

Second, administration capacities are measured by the size of personnel as well as expenditures on information technology in general and in particular, the number of computers. In addition to this, investment in physical capital at government and local institutions are both measured directly.

Third, the transportation infrastructure within counties is captured by the number of telephone lines and the density (i.e. km/area) of various types of road networks (total, motorways, other roads).

Policy variables from the municipality survey

The IEHAS/Median municipality database is composed of two surveys. The first one includes answers to questions on drivers of municipal activities with responses from the Mayor's office. The second survey is filled in by the municipal administration and questions are related to financial features. Below, variables of the vector $local\_avg_{r(t-1)}$ are discussed.

For the basis of this analysis is the county, we simply averaged responses from various settlements within each counties. (The survey included districts of Budapest but unfortunately, excluded Zala county.) This is of course imperfect for several firms are not located in surveyed settlements, and for example, actual local taxes may be quite different from one area to another. Accordingly, insignificant parameters would either signal the lack of explanatory power in economic sense or suggest that settlement level heterogeneity is substantial even within counties that prevents inference.

As for the first survey, the costs of fixed investment is captured by land prices that are given for 1995, 2000 and 2004, so missing years had to be estimated based on these three points using a simple linear method. Prices are related to areas for industrial activity, with utilities and a road connecting the settlement and the area. Not all municipalities gave figures but there were enough to estimate county level averages.

Tax policy is captured both by the nominal tax rates and presence of concessions. As for the taxation variable, the local tax is a turnover tax for companies (i.e. its is based on their output and not the profit). The rate is given by the municipalities ranging between 0% and 2%. The survey included figures for 1992, 1995, 2000 and 2004, so missing years had to be estimated based on these four points using a simple linear method. In addition we have special variables to take into account various concessions offered by municipalities to investors. The first such variable refers to occasions when an area for manufacturing purposes were provided free or with a deep discount for new firms. For every settlement, another dummy takes on 1 where a special tax allowance was promised for new firms (for "recent years") and so it refers to a general approach toward new manufacturing plants. A further concession dummy takes unity when the municipality offers training for new firms.

Finally, a variable is created to measure the importance of large-scale infrastructure related investments between 1995 and 2004. This includes projects defined as "road construction/improvement", "infrastructure development", "transportation development" or "industry parks/areas". The variable ranges between 0 (if there was no infrastructure related investment carried out) and 4 (if four out of five major projects were such investments).

2.4 Results

As for the classic decision variables are concerned, Poisson results (equations [1], [3]) suggest that high per capita income implies more new firms. Of course, when various explanatory variables of development are included in the regression, size and significance of the per capita variable decline. Lower labor costs persistently lead to more new investments as well.

Access variables are important determinants of firm decisions. Substantial production of firms belonging to the same industry is the most stable determinant, but national access to suppliers and customers is also important. The negative sign of the presence of local firms, excluding those operating the same industry, may reflect a strong input competition, while the negative sign of the national presence of firms in the same sector points toward output competition that is not offset by positive externalities. Distance from foreign markets is always one of the strongest determinants of location choice confirming the importance of locating close to export markets. Importantly, the entry of various county feature variables hardly affects the access variables, save the access to business services that is highly correlated with other measures of development.

Looking at the development variables, road network is reassuringly positively related to location choice.[10] Similarly, a positive effect is generated by the development of the telephone network, while a positive but weaker effect is generated by the number of students at local universities. Employment in research and developments centers is also an important factor, while the introduction of other R&D variables provides no significant information any further.

Although the Poisson specification comes from the Random Utility Maximisation framework, the likelihood ratio test of equidispersion fails for all specifications we

[10]When estimated separately, motorways alone enter with a strongly significant coefficient. However, the best explanatory variable is generated by including all types of roads. Results are available on request.

have tried, and the overdispersion parameter ranges mostly between 0.3-0.4. Thus, we turned to the negative binomial specification that allows for overdispersion. Importantly, qualitative results are mostly unchanged (see equations [3] versus [4] or [11] versus [12]) although the significance level would sometimes differ a great deal between the two methods [11]. Note further, that when we include many variables that are correlated with the average wage variable, its the significance would disappear (as in equation [5]). To remedy this, we included an industry specific county wage that is available for 91% of all industry-year-county combinations. Indeed, the wage variable becomes negative and significant once again ([6], [7]).

Apparently, administration capacities (available for 1995-2002 only) matter as the total employment of public administration offices enters strongly suggesting that firms appreciate cities that offer decent administrative services (see equations [6] or [10]). Interestingly, other features such as employment or investment in information technology, seem to have no impact [12].

Higher local taxes are shown to be a deterrent of new firms. In addition to altering taxes, cities can improve business conditions by providing concessions for new firms. The provision of explicit tax allowance has a very strong positive impact and offering education subsidies looks like a decent signal of business friendly environment, too. In contrast, the number of infrastructure related projects seems to be incapable of picking up the pace of development in an area.

The effect of higher land prices in most cases is slightly negative but insignificant - this may be taken as proof that municipalities may give various concessions but leave land prices to market forces. The dummy for special industrial area is very unstable and mostly insignificant. We suspect that a favourable property deal may be offset by signals of a poor area. Remember that these results (equations [4], [8] or [10]-[12]) being based on the municipal survey should be taken with care due to the scarcity of data for several counties.

Finally, we looked at the impact of public investment variables (available for 1996-2002 only) that pick up investment carried out first by the central government and

[11] This robustness is not unusual in the literature, for example Smith & Florida (1994) finds a similar pattern for Poisson, negative binomial and even for the Tobit model. In this paper, we mostly presented results with the negative binomial regression. Note that despite the statistical advantage of the negative binomial model, one may prefer the Poisson given the proximity of actual results and the model's direct link to theory. Results with the Poisson are available on request.

[12] As expected, IT expenditure and number of PCs are closely correlated, and individually both enter with the same sign.

Table 2.1: Do variable groups increase explanatory power? Evidence from LR tests

	Poisson		Neg.Bin	
Group	LR chi2	P-value	LR chi2	P-value
A. ex dummies	250.06	0.000	138.15	0.000
B. ex incentives, taxes	105.68	0.000	48.61	0.000
C. ex development	53.60	0.000	31.07	0.000
D. ex access	566.24	0.000	282.68	0.000
E. ex admin*	31.68	0.000	20.30	0.000

Compared to full model. *administration variables refer to a shorter period (1995-) and are compared to Model B.

second by the local one. It was found that local expenditure is strongly negative while the central government effort is mostly positive but insignificant.[13] This suggests that firms perceive the costs that local investments incur while disregard those in case of central efforts (equation [13]).

To check if policy variables provide addition information on location choice, we run several LR tests, results are shown in Table (2.1). It turns out that each group of variables add some additional explanatory power - both in the Poisson and in the negative binomial models. Overall, these results confirm that all types of variables do indeed influence location choices.

Eventually, location choice features may have changed through time. For several cases, we included time effects to treat some of these problems that may have been masking important effects. Due to the lack of data for several years, a few variables may have lost explanatory power when data is analysed for a sub-period only. For example, the standard deviation of local tax rates between 1993 and 1998 is half than what it is for 1999-2003. When time fixed effects are introduced in equation [14], the negative sign for this variable returns even if being significant at 10% only. There is little difference in terms of significance between various models. We believe that equation [9] seems the best description of the full time period, while [13] is supported for the shorter time period (allowing for the inclusion of more variables).

Overall, we have learn that characteristics of development as well as several municipal and regional decisions, policies can affect location choice. The choice of econometric model has some but little impact, but time dummies remained useful elements

[13]Investment at a regional level is a relatively poor measure and it is biased towards human capital. Thus, it may have little correlation with actual investment in physical capital.

of models.

2.5 Conclusions

In this paper, simple count data models were applied to detect the impact of various factors on location choice of firms. We considered manufacturing companies with foreign ownership setting up a new company in Hungary between 1993 and 2002. Using a set of industry specific access variables with intercompany sales, we found that the proximity to sellers and buyers of potentially important intermediate goods influence location choices. In addition to the location of other firms and wages, it was shown that regional development and some public policy measures will influence decisions. The key variables found here include industry specific wages, output of the actual firm's industry, distance from export markets, density of road network, employment in R&D units. Further, local taxes as well as tax allowance policy of municipalities seem to matter.

It is interesting to compare the effect of variables here and their impact on productivity of existing firms. As shown in Békés & Muraközy (2005), several measures of development proved to be significant in both cases. Most importantly, the density of road network (including motorways) positively influenced location choice and productivity as well. Regarding policy, local investment in public infrastructure has a negative effect in both cases, as firms take investment costs into account, while actions of the central government may have a positive impact especially for the productivity of existing firms. A somewhat larger size of administration helps new firms to settle but later on, it has no effect on productivity. However, the intensity of information technology used in offices contributes positively to corporate TFP. As expected municipal concessions offered for new firms would influence location decisions only.

Our main goal in future research is to analyse the effect of spatial structure in more detail. First, location choice should be considered at the settlement level as well. Second, we are interested to find out if there is a spillover from infrastructure development onto "nearby" regions. Indeed, econometric issues like spatial autocorrelation of development measures should be taken greater care of, too. Third, we plan to look explicitly on the influence of exact proximity to transportation infrastructure (such as motorways) as well as key public institutions (e.g. universities, administration offices).

Table 1 Summary statistics

(to be completed)

Variable	Source	Area	Mean	Std. Dev.
Income per capita	KSH	County	87.234	27.646
Population size	KSH	County	505.85	339.57
ACCESS: local own industry	APEH, AKM	County	649437	2637335
ACCESS: national own industry	APEH, AKM	County	231341	501525
ACCESS: local suppliers	APEH, AKM	County	1050664	2810798
ACCESS: local markets	APEH, AKM	County	1879780	5910528
ACCESS: national suppliers	APEH, AKM	County	354441	574277
ACCESS: national markets	APEH, AKM	County	621667	1041976
ACCESS: business services	APEH, AKM	County	72692050	30586220
LABOUR: average wage	LMS	County	31204.02	14371.81
LABOUR: industry wage	LMS	County	30362.51	16232.25
no. telephone lines	KSH	County /settlement	123244	158637
no. students in university	KSH	County	9803	7332.54
road network density -total	KSH	County	1526.8	563.77
ACCESS: foreign market distance	HAS-IE	County	254	117.478
Road distance between cities	HAS-Institute of Economics	Settlement	190.54	103.01
investment central govt2	KSH	County	18472.36	35035.55
investment local govts2	KSH	County	8964.692	9951.13
R&D employment	KSH	County	2067.28	4706.384
administration employment p.c.	KSH	County	6713.894	14331.58
adminsitration employment in IT p.c.	KSH	County	3281.344	8370.92
administration IT investment p.c.	KSH	County	1362516	6671039
survey local tax rate	KTI Survey	Settlement average	1.202325	.5751337
survey land price	KTI Survey	Settlement average	38.59442	60.09663
survey infrastructure projects	KTI Survey	Settlement average	.8397756	.461991
D(municip. gives special area for investment)	KTI Survey	Settlement average	.7671508	.2032564
D(municip. tax invallowance)	KTI Survey	Settlement average	.6357029	.2238097
D(municip. offers training subsidies)	KTI Survey	Settlement average	.6916895	.2894948

KSH: Hungarian Central Statistics Office, „AKM”: Input-output tables, „LMS”: Annual Labour Market Survey by Ministry of Labour, APEH: Hungarian Tax Authority’s corporate database. NB All variables in estimations are taken in logs.

Table 2. Poisson and negative binomial regressions of location choice

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Estimation method	*poisson*	*neg.bin.*	*poisson*	*neg.bin.*	*neg.bin.*	*neg.bin.*	*neg.bin.*
ln(income p.c.)	0.69***	0.80***	0.76***	0.86***	0.30	0.23	0.32
	(0.15)	(0.19)	(0.15)	(0.19)	(0.24)	(0.24)	(0.27)
ln(county size km2)	0.32	0.02	0.36	-0.00	0.56	0.38	0.37
	(0.25)	(0.31)	(0.27)	(0.35)	(0.44)	(0.43)	(0.46)
ln(LABOUR: average wage)	-0.42***	-0.65***	-0.15	-0.45***	-0.29		
	(0.10)	(0.14)	(0.11)	(0.16)	(0.20)		
ln(LABOUR: industry wage)						-0.78***	-0.85***
						(0.10)	(0.11)
ln(ACCESS: local own industry)	0.24***	0.27***	0.24***	0.27***	0.27***	0.27***	0.26***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
ln(ACCESS: national own industry)	-0.03*	-0.14***	-0.02	-0.13***	-0.12***	-0.11***	-0.11***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
ln(ACCESS: local suppliers)	-0.08***	-0.15***	-0.12***	-0.19***	-0.20***	-0.21***	-0.21***
	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
ln(ACCESS: local markets)	-0.09***	-0.09***	-0.06***	-0.05	-0.08**	-0.04	-0.05
	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)
ln(ACCESS: national suppliers)	0.00	0.21***	0.04	0.25***	0.25***	0.29***	0.29***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)
ln(ACCESS: national markets)	0.16***	0.13***	0.17***	0.14***	0.16***	0.08*	0.09**
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
ACCESS: business services	-0.26***	-0.19**	-0.32***	-0.25***	-0.32***	-0.13	-0.15
	(0.06)	(0.08)	(0.07)	(0.08)	(0.11)	(0.08)	(0.11)
ln(no. telephone lines)	0.15**	0.20**	0.33***	0.38***	0.40**	0.34**	0.24
	(0.06)	(0.09)	(0.07)	(0.10)	(0.16)	(0.16)	(0.17)
ln(no. students in university)	0.68***	0.78**	0.55**	0.71**	0.20	0.15	0.28
	(0.25)	(0.31)	(0.27)	(0.34)	(0.40)	(0.40)	(0.41)
ln(road network density -total)	0.28***	0.24***	0.32***	0.27***	0.21***	0.12**	0.13*
	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)	(0.06)	(0.07)
ln(R&D employment)	0.06**	0.06**	0.09***	0.08***	0.15***	0.13***	0.14***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
ln(survey local tax rate)			-0.25***	-0.22***	-0.08	-0.08	-0.03
			(0.05)	(0.07)	(0.11)	(0.11)	(0.12)
ln(survey land price)			-0.09**	-0.07	-0.07	-0.02	-0.03
			(0.04)	(0.05)	(0.06)	(0.06)	(0.06)
ln(investment central govt2)							
ln(investment local govts2)							
ln (administration employment p.c.)					0.48***	0.43***	0.35**
					(0.12)	(0.12)	(0.14)
ln(adminsitration employment in IT p.c.)					-0.03	-0.03	-0.09
					(0.04)	(0.04)	(0.12)
ln(administration IT investment p.c.)					-0.03	-0.05	0.04
					(0.08)	(0.07)	(0.08)
survey infrastructure projects							
D(municip. gives special area for investment)							
D(municip. tax invallowance)							
D(municip. offers training subsidies)							
ln(ACCESS: foreign market distance)	-0.43***	-0.41***	-0.40***	-0.38***	-0.58***	-0.57***	-0.57***
	(0.05)	(0.07)	(0.05)	(0.07)	(0.10)	(0.10)	(0.10)
FE (time)							*yes*
years included	*1993/2002*	*1993/2002*	*1993/2002*	*1993/2002*	*1995/2002*	*1995/2002*	*1995/2002*
Observations	3000	3000	2850	2850	2280	2091	2091
Model chi-square	7081.70	1874.92	nov.17	1834.67	1409.24	1324.53	1373.82
df	15.00	15.00	17.00	17.00	20.00	20.00	27.00
Log likelihood	-4839.71	-4365.87	-4611.64	-4158.88	-3192.60	-3054.15	-3029.51
Pseudo R2	0.42	0.18	0.43	0.18	0.18	0.18	0.18
overdispersion alpha		0.42		0.41	0.38	0.33	0.30
log likelihood test of alpha=0		947.69		905.52	565.37	493.05	438.14

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3. Poisson and negative binomial regressions of location choice

	[8]	[9]	[10]	[11]	[12]	[13]	[14]
	neg.bin.	*neg.bin.*	*neg.bin.*	*poisson*	*neg.bin.*	*neg.bin.*	*neg.bin.*
ln(income p.c.)	0.03	-0.10	-0.12	-0.31	-0.25	-0.49	-0.40
	(0.24)	(0.25)	(0.28)	(0.25)	(0.30)	(0.31)	(0.32)
ln(county size km2)	-0.33	-0.66*	0.05	-0.04	-0.27	-0.87*	-0.89*
	(0.37)	(0.38)	(0.47)	(0.41)	(0.49)	(0.49)	(0.50)
ln(LABOUR: average wage)			-0.28				
			(0.21)				
ln(LABOUR: industry wage)	-0.72***	-0.84***		-0.81***	-0.86***	-0.72***	-0.80***
	(0.09)	(0.10)		(0.08)	(0.11)	(0.11)	(0.11)
ln(ACCESS: local own industry)	0.26***	0.26***	0.27***	0.27***	0.26***	0.27***	0.26***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ln(ACCESS: national own industry)	-0.12***	-0.11***	-0.12***	-0.03	-0.10***	-0.13***	-0.12***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
ln(ACCESS: local suppliers)	-0.20***	-0.20***	-0.19***	-0.15***	-0.21***	-0.18***	-0.19***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
ln(ACCESS: local markets)	0.04	0.04	-0.04	-0.04	0.01	0.05	0.04
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
ln(ACCESS: national suppliers)	0.29***	0.30***	0.25***	0.10***	0.29***	0.28***	0.29***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
ln(ACCESS: national markets)	0.04	0.05	0.14***	0.10***	0.07	0.00	0.02
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
ACCESS: business services	-0.07	0.10	-0.24**	0.02	0.00	0.06	0.08
	(0.07)	(0.09)	(0.11)	(0.10)	(0.12)	(0.09)	(0.12)
ln(no. telephone lines)	0.27***	0.41***	0.31**	-0.04	0.13	0.33*	-0.00
	(0.10)	(0.12)	(0.16)	(0.13)	(0.18)	(0.18)	(0.23)
ln(no. students in university)	1.06***	0.93**	0.84*	0.89**	0.92**	1.28***	1.58***
	(0.36)	(0.36)	(0.44)	(0.38)	(0.45)	(0.46)	(0.47)
ln(road network density -total)	0.22***	0.09	0.25***	0.16**	0.14*	0.19***	0.22**
	(0.05)	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)	(0.09)
ln(R&D employment)	0.11***	0.10***	0.19***	0.18***	0.19***	0.09*	0.13***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
ln(survey local tax rate)	-0.25***	-0.05	-0.20	-0.18	-0.13	-0.06	-0.34*
	(0.07)	(0.09)	(0.13)	(0.12)	(0.15)	(0.17)	(0.22)
ln(survey land price)	-0.09*	0.02	-0.02	0.06	0.07	0.01	0.01
	(0.05)	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)
ln(investment central govt2)						0.20	0.26
						(0.19)	(0.19)
ln(investment local govts2)						-0.61***	-0.45**
						(0.20)	(0.21)
ln (administration employment p.c.)			0.50***	0.22*	0.31**		
			(0.12)	(0.12)	(0.15)		
ln(adminsitration employment in IT p.c.)			-0.01	-0.02	-0.03		
			(0.04)	(0.09)	(0.12)		
ln(administration IT investment p.c.)			-0.09	0.04	-0.00		
			(0.08)	(0.06)	(0.08)		
survey infrastructure projects	0.08	-0.00	0.07	0.13*	0.10	0.10	0.19*
	(0.08)	(0.08)	(0.09)	(0.08)	(0.09)	(0.10)	(0.11)
D(municip. gives special area for investment)	-0.69***	-0.12	0.16	0.13	0.37	-0.02	-0.14
	(0.25)	(0.29)	(0.33)	(0.30)	(0.35)	(0.32)	(0.37)
D(municip. tax invallowance)	0.57***	0.53***	0.51***	0.54***	0.52***	0.56***	0.62***
	(0.14)	(0.14)	(0.16)	(0.13)	(0.16)	(0.17)	(0.17)
D(municip. offers training subsidies)	1.16***	0.83***	0.59**	0.77***	0.63**	0.71***	0.95***
	(0.22)	(0.23)	(0.27)	(0.23)	(0.27)	(0.26)	(0.28)
ln(ACCESS: foreign market distance)	-0.88***	-0.77***	-0.87***	-0.92***	-0.87***	-0.92***	-0.97***
	(0.11)	(0.11)	(0.13)	(0.11)	(0.13)	(0.14)	(0.15)
FE (time)		*yes*		*yes*	*yes*		yes
years included	*1993/2002*		*1995/2002*	*1995/2002*	*1995/2002*	*1996/2002*	*1996/2002*
Observations	2613	2613	2280	2091	2091	1837	1837
Model chi-square	1757.64	1864.66	1427.20	4638.24	1398.54		
df	21.00	30.00	24.00	31.00	31.00	1135.52	1179.62
Log likelihood	-3959.81	-3906.30	-3183.63	-3231.50	-3017.15	-2647.77	-2625.72
Pseudo R2	0.18	0.19	0.18	0.42	0.19	0.18	0.18
overdispersion alpha	0.35	0.30	0.37		0.29	0.35	0.32
log likelihood test of alpha=0	778.51	670.54	552.05		428.70	445.73	378.05

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Chapter 3

Spillovers from Multinationals to Heterogeneous Domestic Firms: Evidence from Hungary

Joint with Jörn Kleinert[1] and Farid Toubal[2]

Abstract

Firms cluster their economic activities to exploit technological and informational spillovers from other firms. Spillovers through the entry of multinational firms can be particularly beneficial to domestic firms because of their technological superiority. Yet, the importance of foreign firm's spillovers might depend on two key features of domestic firms: their productivity level and its export status. In line with theories and empirical evidence on the absorptive capacity of firms, we argue on the basis of an empirical analysis of Hungarian firms that larger and more productive firms are more able to reap spillovers from multinationals firms than smaller firms. The export status, in contrast, is of minor importance.

3.1 Introduction

The large number of bilateral investment treaties between Central and Eastern European countries and OECD countries agreed on during the nineties as well as various

[1]Department of Economics, Eberhard-Karls-University of Tübingen, Germany

[2]Panthéon-Sorbonne Economics and Paris School of Economics, University of Paris 1, France

investment support schemes carried out since the early nineties suggests that political actors in the participating countries view multinational firms as welfare increasing and growth enhancing. It is widely believed that multinational firms increase competition, transfer technology and help to achieve more efficient allocation of resources. A major argument in this line of reasoning is that inward Foreign Direct Investment (FDI) increases domestic firms' productivity (and thus, enhances economic development) by creating linkages among domestic and foreign firms.

Domestic firms can benefit from the presence of multinationals *in the same* industry through *horizontal* spillovers that might for instance arise through the movement of workers within industries. In addition, there may be *vertical* spillovers from multinationals operating *in other* industries. This type of external effect is usually attributed to buyer-supplier linkages. There are two types of vertical spillovers: backward spillovers are generated through serving customers in downstream industries; forward spillovers are generated through sourcing from upstream industries.

Spillovers from foreign firms are measured through foreign firms' effect on domestic firms' total factor productivity (TFP). The TFP of a firm is the firm-specific component of the firm's technology. A higher TFP of a firm is the result of several factors, such as better use of inputs, more sophisticated sales methods, superior internal organizational structure or simply more knowledge and information. When explaining TFP by spillovers, we make the assumption that the presence of foreign firms creates additional information and opportunities and thereby enhances this firm-specific component of domestic firms' technology. In the literature several channels of positive spillovers have been identified, including labor mobility, supply chains, and face-to-face communication. Yet, while proximity to other producers, customers and suppliers can create a cost advantage or an increase in productivity for a domestic firm, it may also lead to increased competition and to the exit of domestic firms[3].

The empirical literature on FDI spillovers finds mixed support for the positive impact of multinational entry on domestic firms' TFP (Görg & Greenaway (2004)). A large part of literature investigates the extent of horizontal productivity spillovers. Damijan et al. (2003), for instance, use firm level data for several transition countries, including Hungary, and find some evidence for positive spillovers only for Romania. For other countries, the spillover effect is either statistically insignificant or negative. Bosco (2001) analyzes the direct and spillover effects of foreign ownership on firms' TFP in Hungary for the period 1992-1997. She finds that horizontal spillovers are

[3] See Kosova (2006) for a study on the impact of FDI on exit of Czech firms

either insignificant, or negative. According to Aitken & Harrison (1999) and Konings (2001), negative horizontal spillovers arise when multinational firms attract demand away from domestic firms. This lack of sizable horizontal spillovers from multinationals to domestic firms might be explained by the lack of absorptive capacity (i.e. the ability to assimilate and apply new knowledge) of the latter (Girma et al. (2001)). Domestic firms may be unable to learn from multinational firms if the technological gap between the two groups is wide.

Javorcik (2004) extends the spillover approach to backward linkages. Using firm level panel data for Lithuania from 1996 to 2000, she finds evidence of backward linkages. There is, however, no robust evidence from her analysis that domestic firms benefit from horizontal spillovers from multinational firms. Blalock & Gertler (2005) find the same evidence using Indonesian plant-level data. Driffield et al. (2003) examine the relative importance of horizontal, backward and forward spillovers using an industry-level data for UK manufacturing during 1984 - 1992. They show evidence for positive spillovers through forward linkages. There are however no statistically significant effects from horizontal spillovers or from backward linkages.

In this paper, we examine the impact of multinational firms' presence on local firm productivity and size. We assume that the presence of multinational firms generates spillovers which are more important when geographical distance between multinational and domestic firms is small D.B. (1998). For Hungarian firms, this stance is supported by Halpern & Muraközy (2005) who found strong positive spillovers that operate only on small distances (i.e. broadly at the county level) for domestic-owned firms. At the national level, backward spillovers are found significantly positive suggesting that foreign customers make domestically owned firms more productive (Halpern & Muraközy (2005)).

Our aim with this paper is to show how firm level heterogeneity may affect the nature of spillover from the multinational firms to domestic ones. To do this, we run spillover regressions for various types of firms and compare coefficients. We consider two sorts of heterogeneity.

First, we analyze whether more productive and larger firms are able to reap more benefit from spillovers of multinational firms. Many studies find that the degree of technology gap is negatively associated with spillover absorption. For example, Sabirianova et al. (2005) argues that greater technology gap allows for swifter convergence. Similarly, Sjöholm (1999) finds that FDI spillovers are greater in sectors with a high-technology gap in Indonesia. However, the sign may be the other way around, as suggested by the experience of UK establishments, where spillover strength rises in productivity (Girma, Grg 2005). There is also a possibility for hump-shape pattern of the impact, as large gap implies more room for improvement, but also a possibly low capacity to carry out developments.

Our emphasis is on absorptive capacity, a set of organizational routines and processes by which firms acquire, assimilate, transform, and exploit knowledge to produce a dynamic organizational capability. (Zahra & George 2002, p.186.) or in other words, firms ability to recognize valuable new knowledge, integrate it into the firm and use it productively (Cohen & Levinthal (1990)). We argue that identification, acquisition and exploitation may all depend on the firms development, i.e. its level of productivity.

Second, spillovers effect might also differ with respect to the export status of the domestic firm. Exporters' experience in export markets might explain why they deal better with the spillovers of foreign multinational firms (Bernard & Jensen (1999)). However, it might also be that the foreign multinationals' spillovers at home are less important to exporters, because they also learn from firms in the foreign market.

We use a large and extensive data set on Hungarian manufacturing firms. The data set crucially entails information on domestic and export sales as well as ownership. Further, we have information on employment, capital and other firm-level characteristics that enable us to compute the TFP of each domestic firm. We work with an unbalanced panel of manufacturing firms for the period 1992-2003.

Our empirical analysis makes use of three variables which have to be constructed in a first step. (i), we compute the TFP of domestic firms using the semi-parametric Olley & Pakes (1996) methodology. (ii), we construct the horizontal and vertical spillovers variables following Javorcik (2004). We depart from her analysis by taking the extreme view that spillovers from multinationals can only be reaped by domestic firms located in the same county. (iii), we quantify the net effects of spillovers by controlling for the degree of competition. Therefore, we construct a Herfindahl index at sectoral and county level.

We then estimate the effect of multinationals' spillovers on the *average* domestic firm's TFP using a firm fixed-effects panel model. The firm specific effects allow the control of the firm's technology and the isolation of the sectoral spillovers effects. Finally, we are interested in the difference in the effect of spillovers on firms that differ in productivity. We therefore estimate simultaneous quantile regressions.

The remainder of this paper is structured as follow. In Section 2, we provide information on the Hungarian dataset and the descriptive statistics. In Section 3, we present the estimation strategy. In Section 4, we discuss our results. We conclude in Section 5.

3.2 Descriptive Statistics

In this section we present the data and analyze Hungarian firms' productivity. Our analysis is limited to manufacturing firms that meet the data requirements that will be described in the first subsection. In the second subsection, we discuss the distribution of Hungarian firms with respect to size and productivity. As documented for other economies as well, exporters are larger and more productive than domestic firms over the whole size distribution. Foreign multinational firms are larger and more productive than exporters. Hence, it is possible that Hungarian firms (non-exporter and exporter) learn from more productive foreign multinational firms. In the third subsection, we therefore have a first look at our main interest: the relationship of productivity and the number of foreign multinational firms active in a particular Hungarian county.

3.2.1 Data

We use a Hungarian corporate dataset, which is based on annual balance sheet data submitted to APEH, the Hungarian Tax Authority[4]. The dataset contains information on *all* registered, double entry book-keeping firms. The data include the information of a firm's balance sheet and income statement. It entails information on sales, employment, total assets, labor costs, and equity ownership. It also includes information on each firm's sector classification (NACE rev-1, two-digit level) and on the location of the firm's headquarter. The data covers firms' activities between 1992 to 2003.

In Hungary, economic transition has lead to the entry of new domestic and foreign firms. The number of firms has risen substantially from 55,213 in 1992 to 226,072 in 2003. The sample used in this study is less comprehensive than the original APEH data for two reasons. First, we concentrate on manufacturing firms. Second, very small firm data are unreliable and no complete information exists on employment

[4]See details in the Appendix

and fixed assets, which are required to compute the TFP variable. As a result, this sample contains 108,541 observations over 12 years, rising from 6,003 firms in 1992 to 11,208 in 2003. The dataset covers 42% of the total number of manufacturing firms and 73% of total turnover. We use the subsample of domestically-owned firms. It includes 66,470 observations from 11,767 firms for the period from 1993 to 2002.

In table (3.1), summary statistics for all domestically-owned firms in our sample are presented.

Table 3.1: Summary statistics of variables. Domestically-owned firms only

	Mean	Std. Dev.
Fixed assets (log)	8.324	1.967
Sales (log)	10.78	1.547
Materials (log)	9.468	1.579
Employment (log)	2.848	1.242
Domestic Sales (log)	10.80	1.562
Export Sales (log)	9.660	2.357
Export share	0.114	0.249
Exporter status (dum)	0.253	0.435
Horizontal Linkage	0.330	0.224
Backward Linkage	0.145	0.088
Forward Linkage	0.260	0.242
$R\&D$ Linkage	0.119	0.117
Wholesale linkage	0.262	0.192
Herfindahl index	0.137	0.152
Private share	0.974	0.149
TFP (log)	1.815	0.598

3.2.2 Total Factor Productivity, Domestic and International Activities

The data at hand allows discrimination between firms according to their export status and their foreign ownership. We differentiate between four types of firms in the APEH database: domestic non-exporting firms (hence domestic firms), domestic exporters, foreign-owned non-exporting firms and foreign-owned exporters. We use the foreign ownership information to compute our horizontal and vertical spillover variables (see section 3.1) and focus on the impact of multinationals' spillovers on the productivity and size of domestic firms. We define an exporter as a firm that exports at least 5% of its total sales and a foreign owned firm as a firm with at least 10% foreign stake.

In 2002, the sample includes 8,650 domestically owned and 2,112 foreign owned firms. Exporters account for 27% of domestically owned firms and 74.0% of foreign owned firms. The foreign presence in Hungarian manufacturing is rather important, as domestic firms with foreign capital are responsible for 76.6% of total sales in our sample (total sales of foreign firms reached about 28.6 billion euros compared with about 8.7 billion euros by domestically owned ones).

Total factor productivity (TFP) of firms is proxied by an estimated firm-level Solow residual. We use the Olley & Pakes (1996) (OP) semiparametric method to estimate firm-level TFP, a method that takes into account the endogeneity of some inputs, the exit of firms as well as the unobserved permanent differences among firms. We consider the following Cobb-Douglas production function

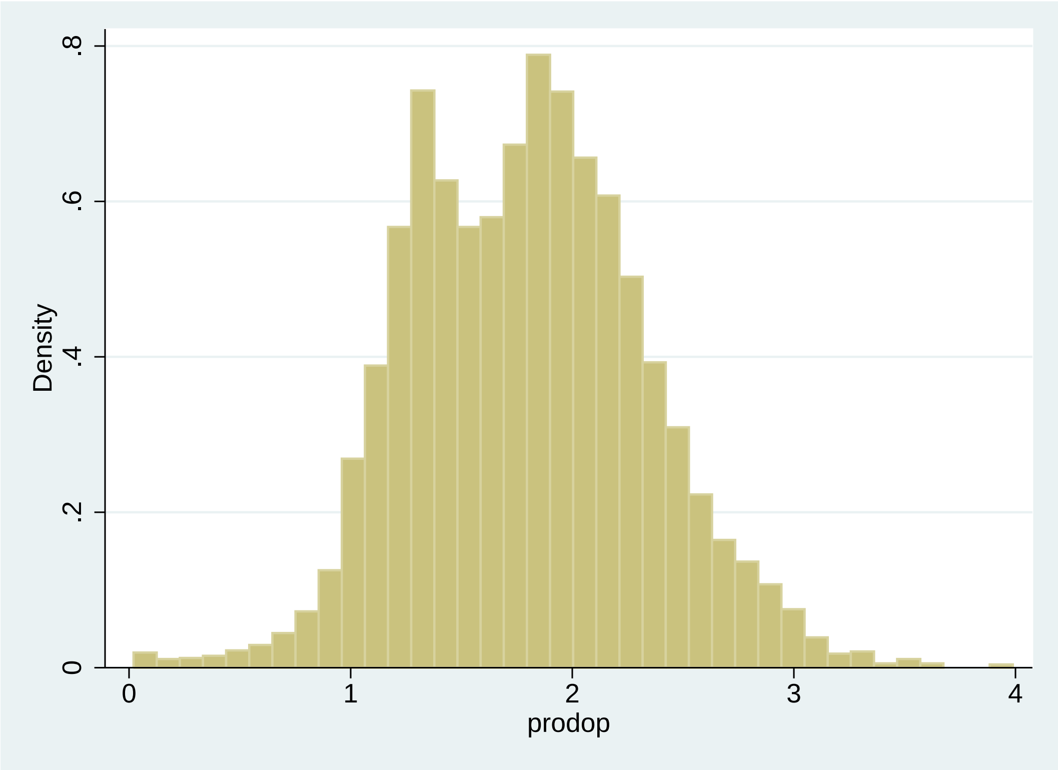
$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + \epsilon_{it} \tag{3.2.1}$$

and denote the logarithm of output (total sales), fixed asset capital, labor (employment) and intermediate inputs (materials) with y_{it}, k_{it}, l_{it} m_{it}, respectively. Subscripts i and t stand for firm and time, ω_{it} denotes productivity, and ϵ_{it} stands for measurement error in output. For details, see the Appendix.

Figure (3.1) shows that the distribution of Hungarian firms' TFP is right skewed. It is, however, not too far from log-normal. We have a closer look at the heterogeneity of Hungarian firms using the results of Table (3.2). We split the distribution of the logarithm of TFP in five intervals and report information on the corresponding number of domestic firms, export status and sales.

Table (3.2) shows two interesting facts. First, the most productive firms are not necessarily the largest with respect to sales. As for both the fifth ([2,3]) and the forth ([3,6.3]) intervals, the share of the interval sales in total sales is below their shares in total number of firms. We expect sector differences behind this finding. Second, export participation increases with productivity. The share of exporters in total firms in the interval increases from 26.2% in the first interval to 41.2% in the fifth. The increase is even more impressive if export activities are measured in export sales instead of number of exporters. Both measures suggest that exporters are more productive than non-exporting domestic firms. The qualitative results of Table (3.2) are robust to change in interval borders.

Figure 3.1: Distribution of Hungarian firms' TFP



Source: APEH, authors' computation.

In Figure (3.2), we show the cumulative distribution of TFP and sales of Hungarian firms according to their export status. Panel (a) of Figure (3.2) points to first-order stochastic dominance of exporters with respect to sales. Exporters are selling more than domestic firms over the whole distribution. The first-order stochastic dominance of exporters with respect to TFP is, however, not obvious from Panel (b) of Figure (3.2).

We use the non-parametric Kolmogorov-Smirnov test (KS-test) to determine whether the sales and TFP distributions between the two groups differ significantly. The KS-test calculates the largest difference between the observed and expected cumulative frequencies, which is called D-statistics. These statistics are compared against the critical D-statistic for the sample size. The results of the two-sided KS-test are shown in Table (3.3).

Concerning the sales distribution, the largest difference between the distribution functions is 0.3034, which is statistically significant at 1%. Thus, the null hypothesis

Table 3.2: Breakdown of Hungarian Firms by Total Factor Productivity in 2000 (percentage into bracket)

$lnTFP_{it}$ interval	Number of firms	Number of exporters	Total Sales	Export Sales
$[-8.2, 0]$	61	16	4.E+06	6.45E+05
	(0.69)	(26.2)	(0.05)	(16.1)
$[0, 1]$	395	107	2.E+08	7.51E+07
	(4.46)	(27.1)	(2.36)	(37.5)
$[1, 2]$	5249	1738	5.E+09	3.07E+09
	(59.26)	(33.1)	(64.14)	(61.4)
$[2, 3]$	2995	1232	3.E+09	1.99E+09
	(33.82)	(41.1)	(32.49)	(66.3)
$[3, 6.3]$	157	65	8.E+07	5.90E+07
	(1.77)	(41.4)	(0.95)	(73.8)

Source: APEH, authors' computation. Sales figures in HUF million.

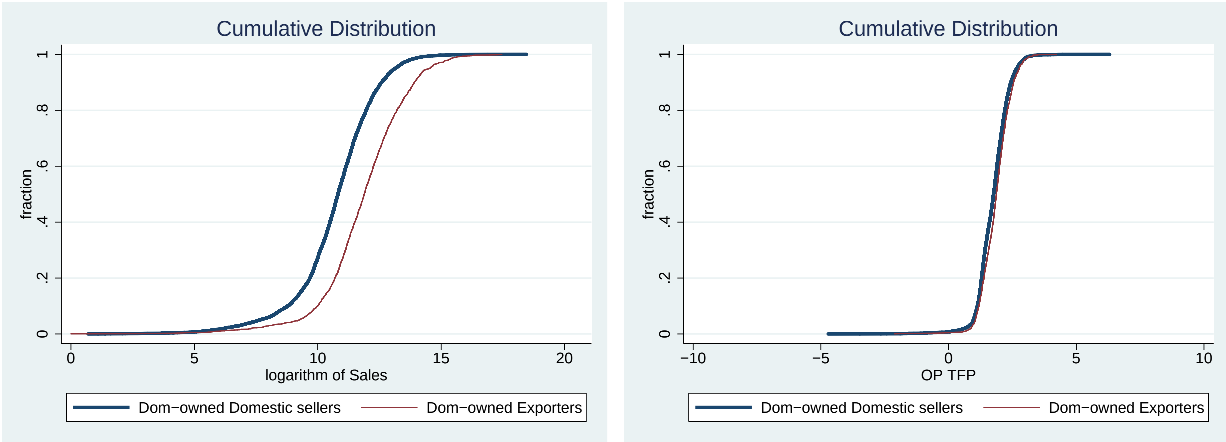
that both sales distributions are equal is rejected. From the left hand-side of the KS-test we can reject the hypothesis that domestic firms are larger than exporters with respect to their sales. The largest difference between the distributions functions is 0.3034, which is statistically significant at 1% level of significance. From the right hand-side of the KS-test, we accept the hypothesis that exporters are larger than domestic firms. The largest difference between the distributions functions is -0.0005, which is not significant. Therefore, we cannot reject the stochastic dominance of exporters' sales distribution over domestic firms' sales distribution. However, we can reject the stochastic dominance of domestic firms' sales distribution over exporters' sales distribution.

We find qualitatively similar results using the TFP distributions. Exporters' TFP cumulative distribution with respect to TFP dominates stochastically domestic firms' TFP cumulative distribution.

As result, the KS-test of stochastic dominance suggests that exporters are more productive than domestic firms and larger in size[5].

[5]Note that the KS-test results are qualitatively similar for each year of the sample.

Figure 3.2: Cumulative Distribution of:



(a) Sales (b) Total Factor Productivity

Source: APEH, authors' computation.

3.2.3 TFP and Spillovers

Having documented that exporters are more productive than domestic firms, we now turn to the most productive firms in Hungary: foreign multinational firms.

Some transition countries, and Hungary in particular, offer a laboratory environment for studying spillover effects as the presence of foreign firms is rather overwhelming. This is true, even if transition started before our sample period of 1992-2003, and foreign firms entered the market as early as the 1989 via joint ventures and greenfield investment. This is how the foreign share in manufacturing sales reached as much as 30% in 1992.

There is substantial sectoral as well as regional disparity in terms of foreign presence. While in *Vas*, a Western county, foreign firms were responsible for two-third of sales in 1992, this share was just over 10% in *Hajdu-Bihar* county, in the South-East of the country. In 2002, almost 90% of manufacturing production was carried out by non-domestic firms, and the lowest share of multinationals in a county rose to half this value (42%-48% in *Bekes*, *Bacs-Kiskun*, *Veszprem*). The picture is similarly diverse in terms of industries. In the production of motor vehicles, non-domestically owned firms were responsible for over 98% (*sic*!) of output in 2002 compared with just over 17% in press and 57% in raw materials and over 65% in machinery and equipment.

We are interested to see whether Hungarian firms (non-exporters and exporters) can learn from foreign multinational firms or use their proximity in another way

Table 3.3: KS-Test of Differences between Exporters and Domestic firms, Sales and TFP, 2000

		Sales	
Group	Largest Difference	P-value	Corrected
$Ho: Exp - Dom \leq 0$	0.3034	0.000	
$Ho: Dom - Exp \leq 0$	-0.0005	0.999	
Combined K-S	0.3034	0.000	0.000
		TFP	
Group	Largest Difference	P-value	Corrected
$Ho: Exp - Dom \leq 0$	0.0918	0.000	
$Ho: Dom - Exp \leq 0$	-0.0014	0.995	
Combined K-S	0.0918	0.000	0.000

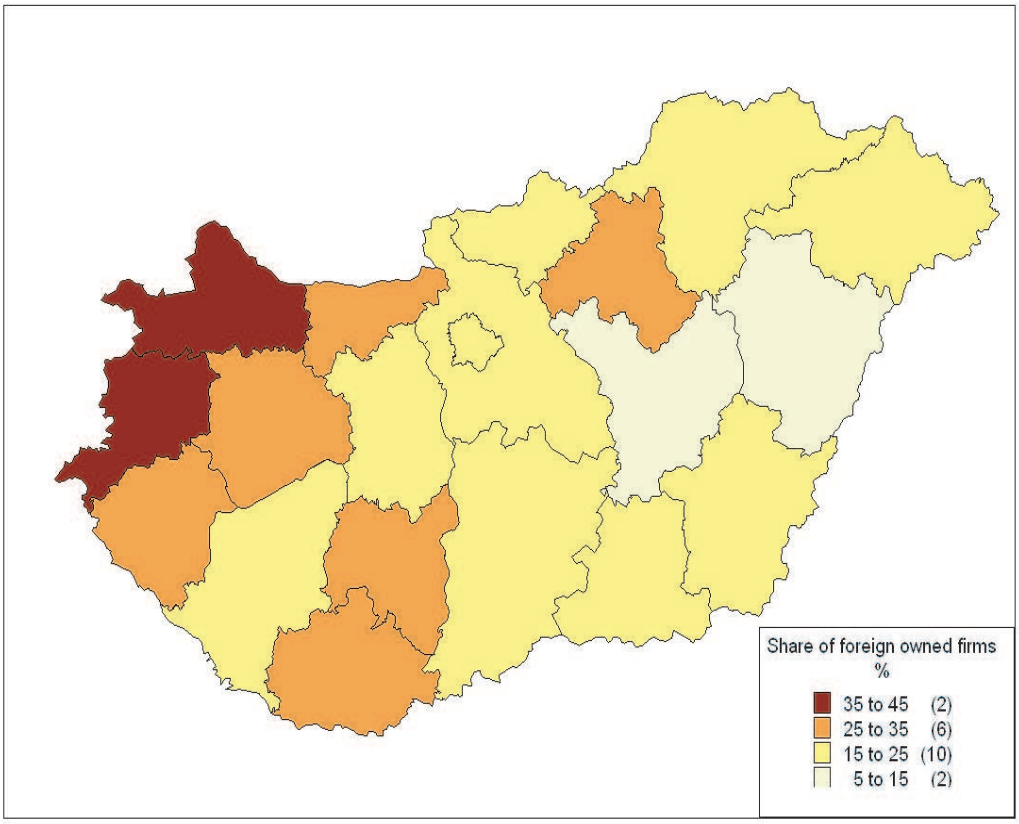
to increase their productivity. We therefore first look at the productivity gap. A productivity gap is the first necessary condition for positive spillovers. Then we look at multinationals' geographic location relative to Hungarian firms. Geographic proximity is the other necessary condition for spillovers.

We use again the KS-test to determine whether the sales and TFP distributions of foreign owned and domestically owned firms differ significantly. We present the comparison of foreign owned firms and the group of Hungarian exporters, which are more productive than Hungarian non-exporters. The results of the two-sided KS-test are shown in Table (3.4). The KS-test reveals that the size of the distribution of foreign multinational firms stochastically dominates those of Hungarian exporters. Thus, the first necessary condition for positive spillovers is met.

Table 3.4: KS-Test of Differences between foreign multinational firms and Hungarian Exporters. TFP, 2000

		TFP	
Group	Largest Difference	P-value	Corrected
$Ho: MNE - Exp \leq 0$	0.0474	0.020	
$Ho: Exp - MNE \leq 0$	-0.0111	0.809	
Combined K-S	0.0474	0.041	0.037

Figure 3.3: Regional distribution of foreign owned firms



Source: APEH, authors' computation, share in percent.

Concerning the second necessary condition, we look at the regional distribution of foreign owned firms. Figure (3.3) shows that Western counties have a higher share of foreign firms, while the Eastern and South-Eastern counties have a rather low share of foreign owned firms.

Next, we look at the relationship of the share of multinational firms in total firms in a particular county and the TFP of Hungarian firms in that county. We regress the logarithm of firm level TFP of domestic firms on the share of multinational firms in sector j of county l, N_{jlt}.

$$TFP_{it} = 0.0692^{***} N_{jlt} + \nu_j + \nu_l + \nu_t \qquad (3.2.2)$$

From this very crude first inspection, we find a positive correlation between a higher share of multinational firms and firm-level TFP. The share of multinational firms and the fixed effect explains 49.67% of the TFP's cross variation.

3.3 Empirical Analysis

In this section, we describe measurement of productivity, detail the spillover variables and give an account of our estimation strategy.

3.3.1 Horizontal and Vertical Spillovers

The total factor productivity of a firm reflects its own technology. Apart from its own technology, the productivity of a firm might also be affected by sectoral linkages and local competition. In this study, we examine the effect of horizontal spillovers, of backward and forward linkages and of local and sectoral competition on firm-specific productivity. Thereby, we describe the logarithm of the TFP of a domestic firm i, in sector j located in a county l at time t, TFP_{ijlt}, as follows

$$TFP_{ijlt} = \alpha H_{jlt} + \beta_1 B_{jlt} + \beta_2 F_{jlt} + \gamma C_{jlt} + \chi Psh_{it} + \nu_i + \nu_j + \nu_t \quad (3.3.1)$$

TFP_{ijlt} has been computed using the semi-parametric estimation suggested by Olley & Pakes (1996). The methodology is developed in Appendix A. It allows to take into account the endogeneity of the inputs in the production function. The endogeneity issue arises because inputs are chosen by a firm based on its productivity.

H_{jlt}, B_{klt}, F_{klt} and C_{jlt} represent local *H*orizontal spillovers, local *B*ackward and *F*orward linkages and local and sectoral *C*ompetition, respectively.[6] We focus on spillovers and competition within a specific county and assume that they arise from the presence of multinational firms in the same county. The variable Psh_{it} stands for the *P*rivatization *sh*are at firm-level (that may change year by year). Since we want to quantify the impact of spillovers at sectoral level on firm-specific total factor productivity, we control for the technology of the firm by introducing firm-specific effects, ν_i. Since the firm specific TFP might also be driven by unobserved sectoral specific shocks, we include a set of sector dummy variables, ν_j. We also assume that firm-specific TFP is affected by macroeconomic shocks and include a set of time

[6]Competition as an influential force on productivity was used e.g. in Nickell (1996).

dummy variables ν_t to control for it. In addition, the time dummy variables control for the average change of productivity that is not due to the spillovers.

Horizontal spillovers occur when entry or presence of multinational firms lead to an increase in productivity of domestic firms active in the *same* industry. This results, for instance, in intra-sectoral movement of workers who take some industry-specific knowledge with them. As in Javorcik (2004), we assume that horizontal spillovers increase with the foreign presence in sector j at time t. We assume, however, that horizontal spillovers are county-specific. We proxy the potential for spillovers by the share of multinational firms in total activities. For each county l, H_{jlt} is defined as foreign equity participation averaged over all firms in the sector, weighted by each firm's share in sectoral output. We proxy horizontal spillovers by H_{jlt} defined as

$$H_{jlt} = \left[\sum_{i \in j,l} share_{it} * Y_{it} \right] / \sum_{i \in j,l} Y_{it} \tag{3.3.2}$$

where $share_{it}$ is the share of firm's total equity that is foreign owned. Y_{it} is the output of firm i at time t.

Vertical spillovers occur when multinational firms' presence in backward or forward industries increases the efficiency of a firm through vertical input-output linkages with suppliers and customers. We calculate the backward linkage with multinational firms (i.e. domestic firms supplying to foreign firms) as

$$B_{jlt} = \sum_{k \neq j,l} \theta_{jk} H_{klt} \tag{3.3.3}$$

where θ_{jk} is the proportion of industry j's output shipped to sector k. This information is taken from the 1998 input-output table at the two-digit NACE level. All results to follow are robust to the use of revised 2000 version. As in Javorcik (2004), the output delivered within the sector is excluded in the computation since this effect is already captured by the horizontal spillovers variable. In other words, the diagonal of the input-output matrix is ignored. By this assumption we acknowledge that the horizontal linkage variable will refer to several sorts of interactions - including the trade otherwise captured by vertical spillover variables. (Of course, a finer sectoral classification would improve upon the problem.)

The forward linkage (i.e. domestic firms purchase goods from foreign firms) is defined as the weighted foreign share in output in the supplying industries.

$$F_{jlt} = \sum_{m \neq j,l} \theta_{mj} H_{mlt} \tag{3.3.4}$$

θ_{jm} is the share of inputs purchased by industry j from industry m in total inputs purchased by industry j. We again exclude the input purchased within the sector because these linkages are captured by the horizontal spillovers variable.

We approximate a potential competition effect by the Herfindahl index. We calculate the Herfindahl indices for all year, sector and county combinations and denote it C_{jlt}. We expect competition to exert a positive effect on TFP. The mode of ownership might also influence the TFP of domestic firms. According to Brown et al. (2006), privately owned firms are more efficient than state-owned firm. We therefore control for the mode of ownership at firm level by including the privatization share.

3.3.2 Estimation Strategy

The heterogeneity in the firm-level data is large. This suggests that we must take it explicitly into account when studying the effects of multinational spillovers on domestic firms. We deal with this large heterogeneity in our empirical analysis in two ways. First, we look at the *average* impact of spillovers and competition on domestic firms. Therefore, we use a firm fixed-effects panel model. While firm heterogeneity is collected in the firm fixed effects, coefficients of H_{jlt}, B_{jlt}, F_{jlt} and C_{jlt} give the average effects of spillovers and competition. Thus, we first ignore differences in the effect of spillovers and competition among firms. Second, we allow spillovers and competition effects to differ between well defined groups of firms but not among firms within each group. We do this by estimating a simultaneous quantile regression model. Unlike the least squares estimator that assumes covariates shifting the location of the conditional distribution only, quantile regression allows us to analyze the possible effects on the shape of the TFP distribution.

In the fixed-effects specification, heteroscedasticity and serial correlation are always potential problems. The bias is larger the longer the time horizon. Since we have short time-series and a large cross-section, it is appropriate to use cluster-sample methods (Wooldridge (2003)) to estimate the fixed-effects model. Cluster-sample methods are a generalization of White's (White (1980)) robust covariance matrices Arellano (1987). The obtained robust variance matrix estimator is valid in the presence of heteroscedasticity and serial correlation provided that, as in our case, T is small compared to the number of groups (Wooldridge (2002*a*), Wooldridge (2003)). The fixed effects panel estimation allow to control for the unobserved domestic firm heterogeneity in the sample. Since our endogenous variable is an estimate itself, we bootstrap the standard errors in a robustness check. This does not alter the significance of the estimated coefficients.

As we have shown in Section 2.2, exporting firms are more productive than non-exporters. That might on the one hand decrease the potential for learning from foreign multinational firms, because more productive firms are already closer to the most efficient technology. On the other hand, learning might be easier because the absorptive capacity of more productive firms is larger. Hence, exporters might be affected differently by foreign multinational firms' spillovers than non-exporting domestic firms. Moreover, there is a second dimension why exporters might reap spillovers to a larger degree: their international experience. Being used to interactions with partners in foreign countries might also ease interaction with foreign multinational firms at home. We therefore test whether spillovers have a different effect on exporters than on non-exporting domestic firms.

The simultaneous quantile regression methodology allows a closer look at the impact of the spillovers on the productivity of domestic firms. We split the firms into twenty groups sorting them with respect to their productivity. We assume firms in each group are affected identically by spillovers and by competition. The bootstrapped variance-covariance matrix takes into account the errors correlation between the different quantiles and allows us to compare coefficients of the explanatory variables in the different quantiles (Koenker & Hallock (2001)). Hence, we test whether spillovers and competition have different impact in different groups. We estimate a simultaneous quantile regression model, which is specified as

$$Quant_{\Theta}\left(TFP_{ijlt}|X_{ijlt}\right) = X^{'}_{ijlt}\beta_{\Theta} \tag{3.3.5}$$

where X_{ijlt} is the vector of independent variables specified in equation (3.3.1) and $Quant_{\Theta}\left(TFP_{ijlt}|X_{it}\right)$ the conditional quantile of TFP. The distribution of the error term ν_{ijlt} is left unspecified so the estimation method is essentially semiparametric. Koenker & Bassett (1978), introducing this technique, show that β_{Θ} can be estimated by

$$min_{\beta}\{\sum_{ijlt:TFP\geq X^{'}\beta}\Theta|TFP_{ijlt}-X^{'}_{ijlt}|+\sum_{ijlt:TFP<X^{'}\beta}(1-\Theta)|TFP_{ijlt}-X^{'}_{ijlt}|\} \tag{3.3.6}$$

The main advantage of the quantile regression approach is that it allows different slope coefficients for different quantiles of the conditional distribution of the TFP variable to be estimated. Since Θ varies from 0 to 1, we trace the entire distribution of TFP conditional on the set of independent variables. As emphasized in Girma et al. (2004), quantile regressions provide a robust alternative to OLS when as in our case the error terms are non-normal. The tests of normality of the TFP distribution, as well as a skewness and kurtosis test, reject the log-normal distribution of TFP. Tests of normality reject a log-normal distribution of establishment-level TFP for any given year and for all domestic-owned firms.[7]

3.4 Results

Discussion of the estimation strategy is now followed by a presentation of main results attained by both fixed effect panel and quantile regressions.

3.4.1 Average Impact of Spillovers on Domestic Productivity

First, we estimate the average impact of the spillover variables on the domestic firm using a firm fixed effects panel model. Since a firm does not change its sector and its county over time, the firm fixed-effects are perfectly collinear with the sector and county fixed-effects. We thus estimate equation (3.3.1) without introducing sector and county fixed-effects. The results are presented in Table (3.5). In the first specification (S1), we show the results of the average spillovers and the competition effect on domestic firms and exporters. In the second specification (S2), the relative average impact of spillovers on TFP with respect to the exporting status of the firm is analyzed. We separate the effect of spillovers from multinational firms on exporters and non-exporting domestic firms by additionally including an interaction term between the spillovers variables and an exporter dummy variable, *Exp*, and an interaction term between the spillovers variables and a non-exporter dummy variable, *Dom*.

Specification (S1) of Table (3.5) shows that the average impact of horizontal spillovers is positive and significant. Therefore, the potential technology transfer

[7] The Shapiro and Francia (Shapiro & Francia (1972)) test, designed for a smaller sample size, yields a p-value of 0.000 to 0.013 for any given year and a p-value of 0.000 for all but two sectors, while the skewness and kurtosis test of D'Agostino et al. (1990) for the whole sample gave a p-value of 0.000.

Table 3.5: Firm-Level Fixed Effects Panel Regression with $lnTFP$ as Dependent Variable:,

	Labels	(S1)	(S2)
Horizontal Spillovers	H_{jlt}	0.0411**	
		(2.41)	
Backward Spillovers	B_{jlt}	-0.0047	
		(0.10)	
Forward Spillovers	F_{jlt}	0.0392	
		(1.38)	
Herfindahl Index	C_{jlt}	-0.0684**	-0.0660**
		(2.41)	(2.34)
Privatization Share	Psh_{it}	0.0660***	0.0660***
		(4.25)	(4.26)
Horizontal Spillovers×Exporter	$H_{jlt} \times Exp$		0.0344
			(1.64)
Backward Spillovers×Exporter	$B_{jlt} \times Exp$		0.1681***
			(2.60)
Forward Spillovers×Exporter	$F_{jlt} \times Exp$		0.0181
			(0.55)
Horizontal Spillovers×Domestic	$H_{jlt} \times Dom$		0.0437**
			(2.36)
Backward Spillovers×Domestic	$B_{jlt} \times Dom$		-0.0545
			(1.10)
Forward Spillovers×Domestic	$F_{jlt} \times dom$		0.0426
			(1.44)
Time Fixed Effects	Yes	Yes	Yes
Observations		66470	66470
Number of groups		11767	11767
R-squared		78.70	79.00

Robust t-statistics in parentheses.
Standard errors have been adjusted for clustering around the firm's identity.
*** denotes statistical significance at one percent level od significance.
** denotes statistical significance at five percent level of significance.
* denotes statistical significance at ten percent level of significance.

from multinationals to domestic firms in the same sector overwhelms the competition effect that arises from the multinational presence. The average impact of forward spillovers is positive but remains statistically insignificant. The coefficient of the backward spillovers variable is very close to zero and insignificant. Both the significant positive effect of horizontal spillovers and the insignificant effect of vertical spillovers differ from Javorcik's results on Lithuanian firms. Turning to the average impact of competition on total factor productivity, we find that a higher Herfindahl index reduces the productivity of domestic firms. Thus, as expected, more competition yields more productive firms. Moreover, as found in Brown et al. (2006), the firm-level privatization share has a positive and significant impact on TFP.

The coefficients of the Herfindahl index and the privatization share variables are robust to the inclusion of the interaction term between the spillover variables and the export status dummy variables (specification (S2) of Table (3.5)). We do not find any statistically significant impact of horizontal spillovers from multinational firms to exporters, while the coefficient of the interaction term between the horizontal spillovers variable and the domestic firms is statistically significant. For backward linkages, the average impact is positive and statistically significant for the exporters only. Thus, while the results for Hungarian exporters are similar to Javorcik's findings. The results for non-exporters, in contrast, differ.

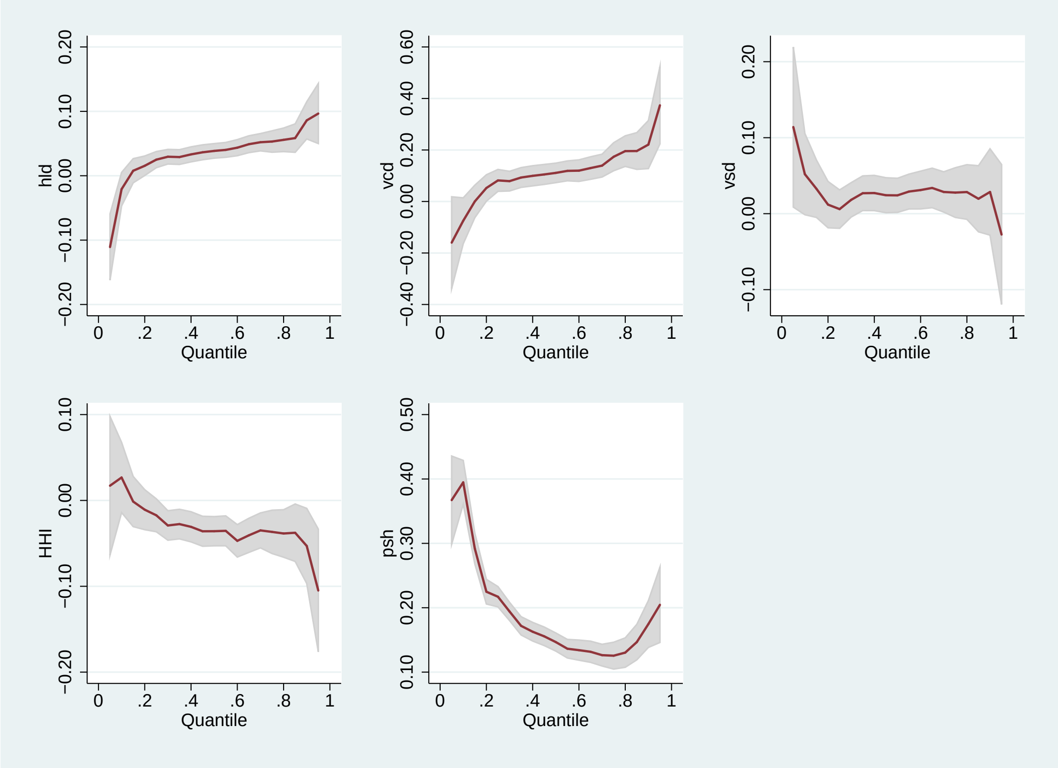
3.4.2 Impact of Spillovers on Heterogenous Domestic Firms

The results of the fixed effects estimation suggest that no vertical spillovers exist from multinational firms to domestic firms. A close look at domestic firm-level heterogeneity might reveal that spillovers from multinationals affect different firms differently depending on their productivity.

We split the distribution of the logarithm of TFP in twenty quantiles and estimate a simultaneous quantile regression. We assume therefore that spillovers and competition effects differ between groups of firms but not within each group. The estimation results are presented in Figure (3.4). In each subfigure, we present the estimated coefficient of each variable on the vertical axis and the corresponding quantile of $lnTFP_{ijlt}$ on the horizontal axis. The first quantile of the distribution contains information on the least productive firms, while the last quantile contains information on the most productive firms.

The results show that horizontal spillovers have a negative impact on the least

Figure 3.4: Simultaneous Quantile Regression: Dependent Variable $lnTFP_{ijlt}$



Estimated coefficient on the vertical axis. Quantile of $lnTFP_{ijlt}$ on the horizontal axis. Source: APEH, authors' computation.

productive firm. This impact is, however, positive and significant for the most productive firm. Moreover, the impact is larger, the more productive is the domestic firm. There is no straightforward explanation, but we suspect there may be two possible reasons for this finding. First, the negative effect on the least productive firm stems from their low level of absorptive capacity. Second, competition from multinational firms, which leads to exit of the least productive firms, stimulate innovation among domestic firms that have high level of productivity (Aghion et al. (2005)). Hence, we argue that the larger the productivity gap between the domestic and foreign firms, the less likely is the domestic firms to gain from foreign multinational firms in its own sector.

We find a negative impact of backward spillovers on the least productive firm, whereas this impact is positive and significant for the more productive firms. The positive impact of backward linkages is increasing with the productivity of the domestic firm. Multinational firms might have a higher incentive to transfer knowledge to more productive firms in their downstream sectors in order to obtain higher sales through higher quality or less expensive goods. Moreover, the increase in foreign presence in the upstream sectors redirects intermediate inputs supply to the downstream sectors away from least productive firms toward more productive firms in the downstream sector. This explains the negative coefficient of backward spillovers on the TFP of least productive firms. The increasing horizontal and backward spillovers with domestic firms' productivity is in line with Girma et al. (2004) findings on UK establishment.

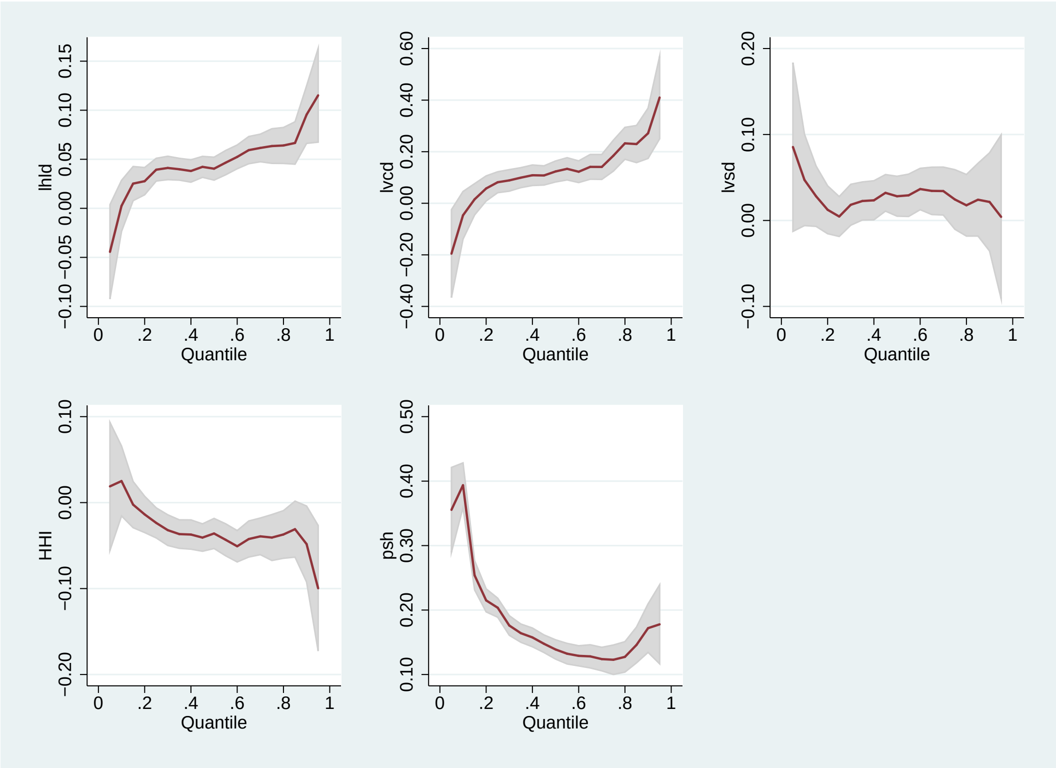
Contrary to Javorcik (2004), we find a positive although small impact of forward spillovers on the productivity of domestic firms. The effect is larger for the least productive firms and insignificant for the most productive firms. The positive effect might stem from a higher quality of inputs purchased from multinational firms.

Turning to the Herfindahl index, it has a positive but insignificant impact on the least productive firms and a negative impact on TFP of more productive firms. Finally, the data suggest a positive correlation between the privatization share and the level of productivity of domestic firms. The impact of privatization is larger the less productive the domestic firm.

Similarly to the location choice problem, it can be argued that spillovers take time to exercise impact. For example, an extension of output by foreign firms may lead to increased interaction with domestic firms at time t, but it is only in time $t+1$ when this relationship bears fruit. Thus, we ran our basic regression with all spillover variables lagged by one year. Results are hardly different. As for the heterogeneity of impact, the "slope" of the backward spillover variable seems slightly stronger now.

As a robustness check, we splat the distribution of the logarithm of TFP in 10 deciles and run fixed effect panel regressions for each deciles. Basic results (presented in Figure (3.6) with a production function and firm-level fixed effects confirm that more productive firms reap greater benefit from backward and to a less extent horizontal spillovers than less productive firms. However, when the OP and firm fixed effects is introduced the image becomes rather blurred.

Figure 3.5: Simultaneous Quantile Regression, lagged spillover: Dependent Variable $lnTFP_{ijlt}$



Estimated coefficient on the vertical axis. Quantile of $lnTFP_{ijlt}$ on the horizontal axis. All spillover variables are lagged one year. Source: APEH, authors' computation.

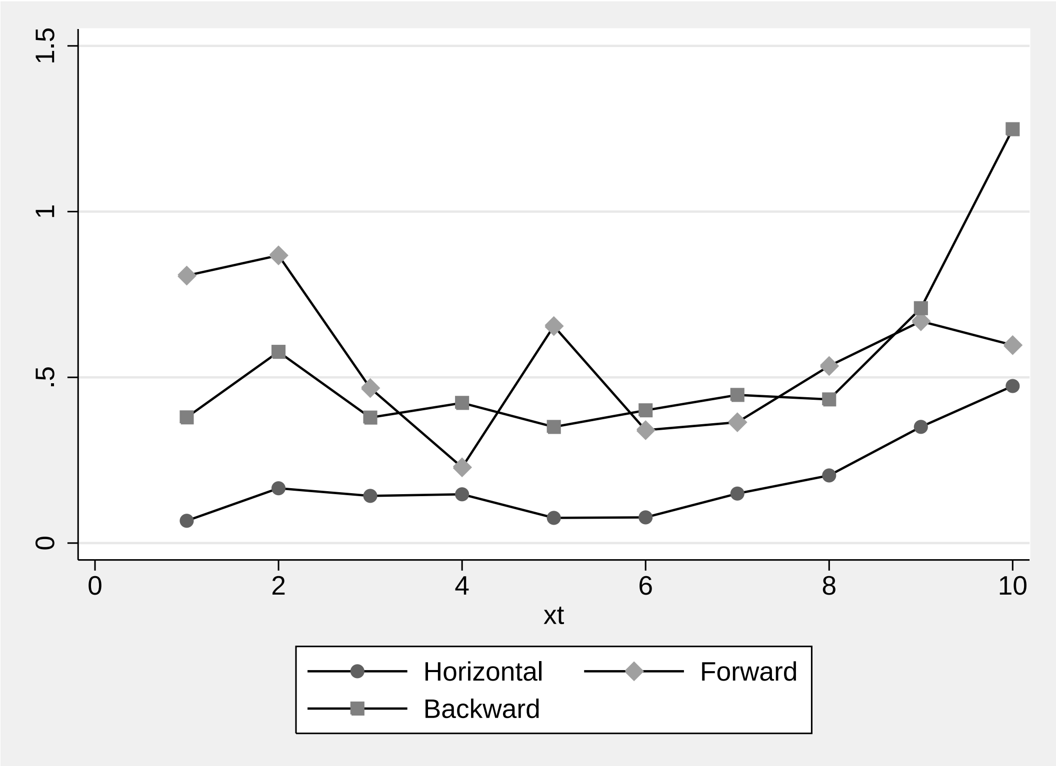
3.4.3 Impact of Spillovers on Exporters and Non-exporters

We separate the effect of spillovers from multinational firms on exporters and non-exporting domestic firms by additionally including an interaction term between the spillovers variables and an exporter dummy variable and an interaction term between the spillovers variables and non-exporter dummy variable.

The results are reported in Figure (3.8). The upper panel of Figure (3.8) show that the coefficients of spillovers from multinational firms to all domestic firms are mainly driven by spillovers to non-exporting firms. Figure (3.4) and the upper panel of Figure (3.8) are very similar. The middle panel shows the coefficients of the spillovers effect on exporters. The bottom panel shows the coefficients of Herfindahl index and of the privatization share variables.

We can statistically distinguish the impact of spillovers from multinational firms by

Figure 3.6: Fixed effect panel regression by deciles: Dependent Variable $lnTFP_{ijlt}$



Estimated coefficient on the vertical axis. Quantile of $lnTFP_{ijlt}$ on the horizontal axis. Source: APEH, authors' computation.

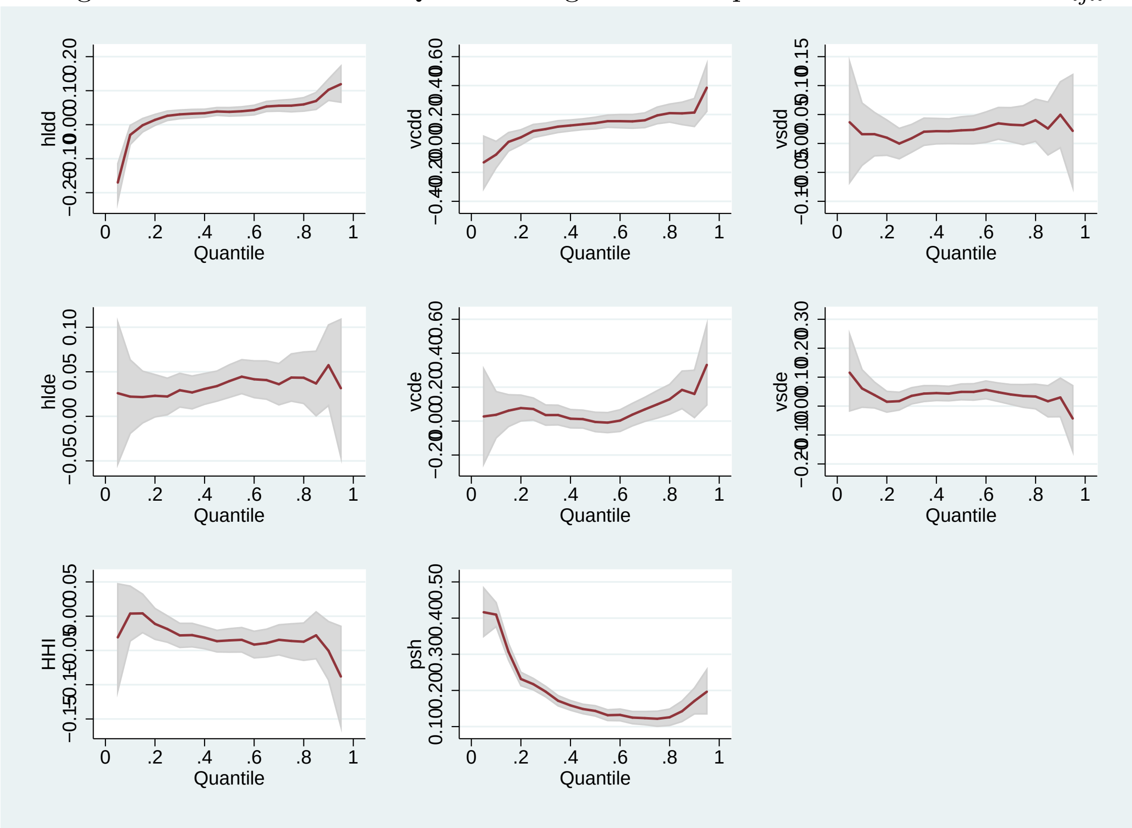
the export status of domestic firms for some quantiles. Most non-exporting Hungarian firms receive horizontal spillovers from multinational firms. The effect of spillovers on TFP increases in productivity.

As for backward linkages, non-exporters gain from positive spillovers if their productivity places them at least in the third decile. The exporters pattern has a slight hump shape, but significant gain from productivity takes place in the upper third of the distribution only. Forward spillovers are very similar for the two categories, slightly positive or zero, for both groups.

The productivity advantage of exporters which we reported in Section 2 therefore does not result from higher spillovers that exporters as such receive from multinational firms relative to non-exporters.

In line with the results from the fixed effects regression, the quantile regressions revealed no larger spillovers for exporters than for non-exporting domestic firms.

Figure 3.7: Simultaneous Quantile Regression: Dependent Variable $lnTFP_{ijlt}$



Estimated coefficient on the vertical axis. Quantile of $lnTFP_{ijlt}$ on the horizontal axis. Source: APEH, authors' computation.

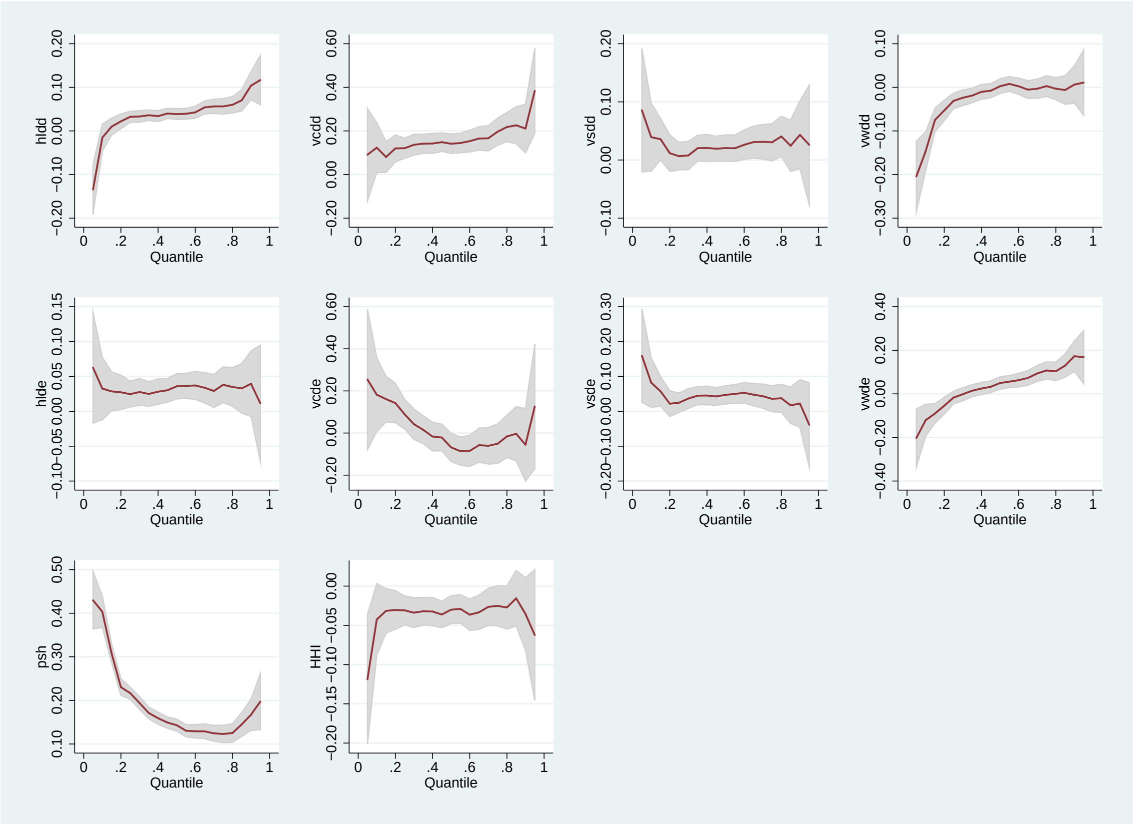
Hence, larger spillovers from multinational firms are not part of the explanation why exporters have higher total factor productivity. Thus, while exporters might receive additional spillovers in the foreign market which increases their TFP, we did not find support for higher spillovers received by exporters at home.

There are three explanations for these findings. First and probably most important, the higher TFP of exporting firms relative to non-exporters is explained by the fact that more productive firms self-select into exporting (as in Melitz (2004)) Thus, exporting status *per se* gives no reason for a difference in impact. Second, exporters might receive additional spillovers in the foreign market which increase their TFP. Third, exporters might learn from foreign owned firms active in the Hungarian wholesale sector because they share a common "*trade technology*".

For the first two points have been examined in literature, we now test the validity of the third assertion by looking at the impact of the share of foreign-owned firms in the Hungarian wholesale sector. Therefore, we construct a wholesale spillover variable, W_{jlt}, that is the share of foreign ownership among firms that operate in the wholesale sector and are exporters:

$$W_{jlt} = \left[\sum_{i \in j, exp=1, l} share_{it} * Y_{it} \right] / \sum_{i \in j, exp=1, l} Y_{it} \qquad (3.4.1)$$

Figure 3.8: Simultaneous Quantile Regression: Dependent Variable $lnTFP_{ijlt}$



Estimated coefficient on the vertical axis. Quantile of $lnTFP_{ijlt}$ on the horizontal axis. Source: APEH, authors' computation.

We find a strong negative impact of foreign-owned importers on the least productive domestic firms whereas this impact is positive and significant for most exporters. While domestic firms might suffer from import competition, exporters might benefit from foreign-owned importers' trade knowledge.

3.5 Conclusions

We examined the impact of the presence of foreign multinational firms in local Hungarian markets on Hungarian firms' productivity. We searched for horizontal spillovers from multinational firms in the same sector, backward spillovers from multinationals that are customers of Hungarian firms and forward spillovers from multinationals that are input suppliers. We used a sample of $11,767$ Hungarian firms and their activities between 1993 and 2002. For this sample, we found significant horizontal spillovers in a firm level fixed effect regression but no evidence of backward and forward spillovers.

Yet, the spillover effects are average effects over all firms which might not be very informative if Hungarian firms are very heterogeneous and this heterogeneity affects the size of the spillovers. We documented great heterogeneity among Hungarian firms with respect to their productivity and size and analyze whether more productive and larger firms are able to reap more benefit from spillovers of multinational firms than less productive smaller firms. We used simultaneous quantile regression to analyze group specific effects with groups defined with respect to productivity. We found significant differences among the groups with more productive firms receiving more horizontal and backward spillovers from foreign multinational firms but less forward spillovers than less productive firms.

There is a second obvious characteristic in which firms differ: their export status. Export status is not independent of productivity since only more productive firms generate profits in the export market. We expected export status to have an effect for two reasons. First, as argued above, exporters are more productive. That might increase the spillovers reaped since the absorptive capacity is larger or decrease the spillover effect because the gap to the most efficient firm is smaller. Second, exporters are used to interact with foreign firms and therefore able to gain more from the presence of foreign multinational firms in Hungary. In a fixed effects regression which separates the spillover effects on exporters and non-exporters, we found significantly positive backward spillovers of multinational firms on Hungarian exporters but no effect on Hungarian non-exporters. Horizontal spillovers in contrast were only significant for non-exporting firms. In line with the results from the fixed effects regression, the quantile regressions revealed no larger spillovers for exporters than for non-exporting domestic firms.

Overall, we found that heterogeneity in terms of productivity influences domestically owned firms' capacity to absorb knowledge and achieve higher productivity.

This finding may have policy implications regarding FDI subsidies, a point left for future research. Another point for further research is related to the potential of omitted variable bias. We understand that it is rather difficult to disentangle several coinciding developments of transition: reallocation of resources, privatization, capital and knowledge inflow or institutional reform. Some of the factors are controlled for via dummy variables, but many must have been left outside of the analysis, potentially affecting spillover.

Chapter 4

Appendix

4.1 Firms versus plants

A key issue is the exact nature of firm location. In effect, most Krugman-type new economic geogrpahy models, as in Fujita et al. (1999), assume that single-plant companies serve the market (each producing a differentiated good). Thus, plant level data would be necessary to represent the actual production site. However, only firm-level data are available instead. As a result, we may have data on a corporate headquarter, rather than its production plant distorting our results a great deal. To check this, two exercises were carried out.

First, the National Corporate Register was consulted to see how large foreign manufacturers such as Siemens, Philips or IBM were incorporated in Hungary. Apparently, these multinationals established separate entities for many of their operations. Siemens AG, a German electronics good manufacturer established a dozen firms up to 2003 including Siemens kft, responsible for all retail activities, Siemens "Finance" (Financial Services), or Siemens "Telefongyár" (Telecom). IBM has its main production plant as part of IBM "Data Storage Systems" in Székesfehérvár (Fejér county), while consulting business is carried out via IBM "Üzleti Tanácsadó" registered in Budapest downtown.

The best example for separation of plants by industries may be the Dutch giant, Philips. Table 6. gives a full account of its presence in Hungary as of 2002. There are nine entities owned by Philips (via one of its entities: Philips Communications

& Processing Services, Philips Beteiligungs GmbH, Koninklijke Philips Electronics N.V.) and most of them has one address only. Some, especially those with trading activities would have two locations. The multinational organisation has invested in various firms including Philips "Components" (machinery) in Győr (Győr-Sopron-Moson county), Philips "Industries Hungary" (electronics) in Székesfehérvár (Fejér county), or Philips "Hungary trading" in Budapest. All Hungarian outlets may be found in one of two Western counties with services and trading activies located in Budapest. A similar structure may be perceived by many other major multinational companies including Audi producer "Porsche Inter Auto", or Electrolux, whose production plant is situated somewhere in the countryside as one firm, while another one in Budapest is responsible for sales or foreign trade.

One should expect that the most problematic bias would come from an over-representation of the capital city given that many firms that entered Hungary, first established a HQ in Budapest. Thus, in a second effort, industry-level aggregates from two sources were compared: The APEH complete firm-level corporate dataset and plant level employer data of the Labour Market Surveys. It showed that the share of Budapest by industries is just a few percentage points higher in the firm-level data. This also supports the assumption that the application of firm-level data should be of no great concern in our practice.

4.2 Corrections to the data - Chapters 2 & 3

There has been serious effort invested in cleaning the data and several corrections were made to the original APEH dataset by the Magyar Nemzeti Bank, the CEU Labour Project[1] and the author. Three important steps have been taken.

First, longitudinal links for foreign firms were improved using data provided by Hungarian statistics office KSH on corporate entry and exit. CEU Labour Project looked for other longitudinal links in which the firms did not simply appear under a new id number, but actually split up into several firms or were formed via a merger. These allowed to keep track most but not all of firms under transformation.

Second, the ownership structure of new firms was repaired in many cases to make sure that foreign ownership reflected the most likely case. Information from balance sheets and adjacent years' values were used.

[1]For a description, see Telegdy (2004). Details are available from the author on request.

Third, sales data for all firms were checked to avoid typing errors. For many firms, sales data were missing. Further problems I found and/or learned from others working with the same or similar datasets included: (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. Overall, I made modifications reaching almost 2% of the total dataset. In some cases, sales could be estimated by using other balance sheet figures, and in others, the simple average of sales data at $(t-1)$ and $(t+1)$ was used.

4.3 Data description of Chapter 4

First note, that the APEH data provide information of firms with non-consolidated accounts. Thus, a manufacturing firm can mostly be considered as an establishment: i.e. a headquarter and a plant.

This version of the dataset comes from the Central European University - Labor Project and is based on a dataset managed by the Magyar Nemzeti Bank. Several steps have been made to improve the consistency of the dataset. The initial dataset were exhaustively cleaned by the CEU Labor Project and the authors.

Non-surprisingly in a transition economy, firms frequently changed their attributes. First, we had to define manufacturing firms and their sector classification to avoid firms appearing/disappearing based on their statistical status. A sector was defined based on the NACE 2-digit code a firm most often used. A firm was kept in manufacturing if it spent 75% of its time in the sample as a manufacturer. Second, longitudinal links for firms had to be improved using data provided by Hungarian statistics office KSH on corporate entry and exit. These are cases, when a firm changes its identification code but remains basically the same. This is especially frequent phenomenon in transition economies such as Hungary, see Brown et al. (2006). Other longitudinal links were investigated where firms did not simply appear under a new code but actually split up into several firms or were formed via a merger. These allowed keeping track of most but not all of firms under transformation. Further, small firms (ones that never had as many as 5 employees) had to be dropped for the well-documented lack of reliable data (see Katay & Wolf (2006)) We discarded 58% of firms for missing or unreliable data. Otherwise, no outliers were dropped.

We made several fixes, too. Obvious typing errors were corrected. In order to ensure that small firms are not dropped for missing data in employment or fixed assets, for missing years we replaced these variables with the mean of their (t-1) and (t+1) values. This was the case for 1175 occasions for employment and 206 cases for fixed assets. Ownership also had to be cleaned for the large number of missing observations (filled in case of equality of the (t-1) and (t+1) values) and typos.

The capital variable was created and corrected following suggestions in Katay & Wolf (2004). Importantly, capital was recalculated by the perpetual inventory method (PIM). The reason for this is that capital stock should be registered at market prices. This is not the case in Hungary, where the stock enters the balance sheet on the book value. Without information on the composition of the capital, actual data represents a mixture of various kinds of assets in terms of age and readiness to use. Hence, the need to recompose the capital stock by the PIM using an initial condition (i.e. first year of investment) and a capital accumulation equation to reconstruct the stock of capital. As a result, investments are deflated by the investment price deflator, and then, the rate of depreciation is used to get K, the capital stock. Thus:

$$K_{i,t} = K_{i,t-1} * (1 - Depreciation_{i,t}) + Investment_{i,t} \tag{4.3.1}$$

Description of variables are presented in table 4.1.

4.4 TFP Measurement methodology in Chapter 4

We use the Olley & Pakes (1996) (OP) semiparametric method to estimate firm-level TFP. This method allows robust estimation of the production function. It takes into account the endogeneity of some inputs, the exit of firms as well as the unobserved permanent differences among firms. The main assumption the OP technique relies on, is the existence of a monotonic relationship between investment and firm-level unobserved heterogeneity. Table (4.2) gives an account of estimated coefficients.

We consider the following Cobb-Douglas production function

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + \epsilon_{it} \tag{4.4.1}$$

and denote the logarithm of output, capital, labor and intermediate inputs with y_{it}, k_{it}, l_{it} m_{it}, respectively. Subscripts i and t stand for firm and time, ω_{it} denotes

Table 4.1: Description of variables

Variable	Details	Source
Output	Net sales by the firm, deflated by sectoral PPI deflators	APEH:income statements
Capital	Fixed assets capital generated and corrected by the perpetual inventory method, following suggestions in Katay and Wolf (2004, 2006)	APEH: income statements
PPI	Producer price deflator, sectoral level	KSH
Ownership	Foreign-owned firms: at least 10% of equity capital is owned by non-residents. (NB. Distribution of the status is bimodal, and results are insensitive to the threshold.)	APEH:balance sheets
Private share	Share of equity capital owned privately (i.e. non-state and non-municipal owners	APEH: balance sheets
Export status	Exporter firm is defined if net export sales reached at least 5% of total net sales. (NB. Distribution of the status is bimodal, and results are insensitive to the threshold.)	APEH:income statements
Investments	Change in fixed assets, reduced by a sector specific depreciation rate calculated from the data, deflated by investment input prices. (NB. Results robust to flat depreciation rate)	APEH: income statements
Investment price deflator	Estimated by authors based on 80% machinery and 20% property price deflators	KSH, authors
Depreciation rate	Directly is estimated from the APEH data. To see robustness of the APEH data, an average of 20% was used, without sizeable impact	authors calc.
Labor	Average annual employment in the given year	APEH:income statements
Materials	All materials, calculated following Katay-Wolf (2006) who advised on how to take care of changes in the accounting law in 2001.	APEH:income statements

productivity, and ϵ_{it} stands for measurement error in output. It is assumed that ω_{it} follow an exogenous first order Markov process:

$$\omega_{it+1} = E[\omega_{it+1}|\omega_t] + \eta_{it+1} \tag{4.4.2}$$

where η_{it} is uncorrelated with the productivity shock. The endogeneity problem stems from the fact that k_{it} and l_{it} are correlated with the ω_{it}. This makes β_{OLS} to be biased and inconsistent. Given that investment is strictly monotonic, it can be inverted as:

$$\omega_{it} = h(i_{it}, k_{it}) \tag{4.4.3}$$

and substituting this function in the production function leads to

$$y_{it} = \beta_l l_{it} + \beta_m m_{it} + \Phi(i_{it}, k_{it}) + \epsilon_{it} \tag{4.4.4}$$

where $\Phi(i_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + h(i_{it}, k_{it})$. Since the functional form of $\Phi(\cdot)$ is not known, we cannot estimate the coefficients of the capital and labor variable directly. Instead, we use a linear model that includes a series estimator using a full interaction term polynomial in capital and investment to approximate $\Phi(\cdot)$. From this first stage, the consistent estimates of the coefficients on labor and material inputs as well as the estimate of the polynomial in i_{it} and k_{it} are obtained.

The second stage takes into account the survival of firms. These probabilities are given by

$$\begin{aligned} Pr\{\chi_{t+1} = 1|\underline{\omega}_{t+1}(k_{t+1}), Jt\} &= Pr\{w_{t+1} \geq \underline{\omega}_{t+1}(k_{t+1})|\underline{\omega}_{t+1}(k_{t+1}), \omega_t\} \\ &= \varphi\{\underline{\omega}_{t+1}(k_{t+1}), \omega_t\} \\ &= \varphi(i_t, k_t) \\ &= P_t \end{aligned} \tag{4.4.5}$$

The probability that a firm survives at time $t+1$ conditional on its information set at time t, J_t and ω_{t+1}. This is equal to the probability that the firm's productivity is greater than a threshold,$\underline{\omega}_{t+1}$, which in turn depends on the capital stock. The survival probability can be written as a function of investment and capital stock at time t. Thus, we estimate a probit regression on a polynomial in investment and capital controlling for year specific effects. Now, consider the expectation $y_{t+1} - \beta_l l_{t+1}$ conditional on the information at time t and survival at $t+1$.

$$\begin{aligned} E[y_{t+1} - \beta_l l_{t+1} | k_{t+1}, \chi_{t+1} = 1] &= \beta_0 + \beta_k k_{t+1} + E[\omega_{t+1} | \omega_t, \chi_{t+1} = 1] \quad (4.4.6) \\ &= \beta_k k_{t+1} + g(\underline{\omega}_{t+1}, \omega_t) \end{aligned}$$

ω_{it} follow an exogenous first order Markov process. We substitute the productivity shock in the above equation using the result from the first stage.

$$y_{t+1} - \beta_l l_{t+1} = \beta_k k_{t+1} + g(P_t, \Phi_t - \beta_k k_t) + \eta_{t+1} + \epsilon_{it} \quad (4.4.7)$$

The third step takes the estimates from β_l, Φ_t, and P_t and substitutes them for the true values. The series estimator is obtained by running a non-linear least squares on the equation

$$y_{t+1} - \beta_l l_{t+1} - \beta_m m_{t+1} = c + \beta_k k_{t+1} + \sum_{j=0}^{s-m} \sum_{m=0}^{s} \beta_{mj} (\hat{\phi}_t - \beta_k k_t)^m \hat{P}_t^j + e_t \quad (4.4.8)$$

where s is the order of the polynomial used to estimate the coefficient on capital.

Table 4.2: Productivity function coefficients

Sector	Observations	Labor	Materials	Capital	Scale
17	4883	0.32	0.66	0.07	1.05
18	6526	0.45	0.59	0.05	1.09
19	2857	0.41	0.57	0.07	1.04
20	6209	0.19	0.82	0.03	1.04
21	1523	0.15	0.83	(0.02)	0.99
22	9010	0.18	0.80	0.06	1.04
24	2978	0.14	0.86	0.03	1.02
25	6097	0.20	0.78	0.06	1.04
26	4100	0.21	0.79	0.07	1.07
27	1501	0.13	0.83	0.06	1.02
28	15736	0.26	0.73	0.05	1.04
29	12104	0.26	0.73	0.02	1.01
30	662	0.42	0.60	0.17	1.19
31	3369	0.27	0.70	0.09	1.06
32	2568	0.27	0.74	0.07	1.08
33	3571	0.26	0.78	0.05	1.09
34	1466	0.28	0.75	0.02	1.05
35	558	0.35	0.78	(0.00)	1.13
36	5762	0.26	0.72	0.07	1.04

NB Figures in brackets are not significant at one percent level of significance.

Bibliography

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & Howitt, P. (2005), 'Competition and innovation: An inverted u relationship', *Quarterly Journal of Economics* pp. 701–728.

Aitken, B. & Harrison, A. (1999), 'Do domestic firms benefit from direct foreign investment? evidence from venezuela?', *American Economic Review* **89**, 605–18.

Amiti, M. & Javorcik, B. S. (2003), Trade costs and location of foreign firms in china. World Bank.

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

Arellano, M. (1987), 'Computing robust standard errors for within-groups estimators', *Oxford Bulletin of Economics and Statistics* **49**(4), 431–34.

Baldwin, R. E. (2004), Industry location: the causes. Graduate Institute of International Studies, Geneva.

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

Barrios, S., Strobl, E. & Gorg, H. (2003), Multinationals' location choice, agglomeration economies and public incentives, Discussion Paper 17, CORE.

Barta, G. (2003), Developments in the geography of hungarian manufacturing. in Munkaerõpiaci Tükör.

Bartik, T. J. (1985), '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,* **3**(1), 14–22.

Basile, R. (2004), 'Acquisition versus greenfield investments: the location of foreign manufacturers in italy', *Regional Science and Urban Economics* **34**(1), 3–25.

Békés, G. (2005), Location of manufacturing fdi in hungary: How important are inter-company relationships?, Working Paper 2005/7, Magyar Nemzeti Bank.

Békés, G. & Muraközy, B. (2005), Firm behavior and municipal policy : The case of hungary, Discussion Paper 05/04, MTA KTI.

Bernard, A. B. & Jensen, J. B. (1999), 'Exceptional exporter performance: Cause, effect, or both?', *Journal of International Economics,* **47**(1), 1–25.

Blalock, G. & Gertler, J. (2005), How firm capabilities affect who benefits from foreign technology. Cornell University.

Bosco, M. (2001), 'Does fdi contribute to technological spillovers and growth? a panel data analyis of hungarian firms', *Transnational Corporations* **10**, 43–68.

Boudier-Bensebaa, F. (2005), 'Agglomeration economies and location choice: Foreign direct investment in hungary', *Economics of Transition* **13**(4), 605–628.

Brown, D., Earle, J. & Telegdy, Á. (2006), 'The productivity effects of privatization: Longitudinal estimates from hungary, romania, russia, and ukraine', *Journal of Political Economy* **114**(1), 61–99.

Cameron, C. A. & Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Cambridge University Press.

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.

Cohen, W. & Levinthal, D. (1990), 'Absorptive capacity: A new perspective on learning and innovation', *Administrative Science Quarterly* **35**, 128–152.

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

Crozet, M., Mayer, T. & Mucchielli, J.-L. (2004), 'How do firms agglomerate? a study of fdi in france', *Regional Science and Urban Economics* **34**(1), 27–54.

D'Agostino, R., Balanger, A. & Jr., R. D. (1990), 'A suggestion for using powerful and informative tests of normality', *American Statistician* **44**, 316–321.

Damijan, J., Knell, M., Majcen, B. & Rojec, M. (2003), 'The role of fdi, r&d accumulation and trade in transferring technology to transition countries', *Economic Systems* **27**, 189–204.

D.B., A. (1998), 'Agglomeration and the location of innovative activity', *Oxford Review of Economic Policy* **12**(2), 18–29.

Deichmann, U., Kaiser, K., Lall, S. V. & Shalizi, Z. (2005), Agglomeration, transport, and regional development in indonesia, Policy Research Working Paper 3477, The World Bank, Washington D.C.

Devereux, M. P., Griffith, R. & Simpson, H. (2003), Agglomeration, regional grants and firm location. University of Warwick.

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

Driffield, N., Munday, M. & Roberts, A. (2003), 'Foreign direct investment, transactions linkages, and the performance of the domestic sector', *International Journal of the Economics of Business* **9**(3), 335–351.

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, KTI.

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

Figueiredo, O., Guimaraes, P. & Woodward, D. (2002*b*), Modeling industrial location decision in us counties. University of South California.

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

Freidman, J., Gerlowski, D. & Silberman, J. (1992), 'What attract foreign multinational corporations? evidence from branch plant location in the united states', *Journal of Regional Science* **32**, 403–418.

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

Gabe, T. M. (2003), 'Local fiscal policy and establishment growth', *Journal of Regional Analysis and Policy* **33**(1), 57–80.

Girma, S., Greenaway, D. & Kneller, R. (2004), 'Does exporting increase productivity? a microeconometric analysis of matched firms', *Review of International Economics* **12**(5), 855–866.

Girma, S., Greenaway, D. & Wakelin, K. (2001), 'Who benefits from foreign direct investment in the uk?', *Scottish Journal of Political Economy* **48**, 119–133.

Görg, H. & Greenaway, D. (2004), 'Much ado about nothing? do domestic firms really benefit from foreign direct investment?', *World Bank Research Observer,Oxford University Press* **19**(2), 171–197.

Halpern, L. & Muraközy, B. (2005), Does distance matter in spillover?, Discussion Paper 4857, CEPR.

Hausman, J. & McFadden, D. (1984), 'A specification test for the multinomial logit model', *Econometrica* **52**, 1219–1240.

Head, K. & Mayer, T. (2004), The empirics of agglomeration and trade, *in* V. Henderson & J. F. Thisse, eds, 'Handbook of Regional and Urban Economics, Vol 4.', number 15, North-Holland, Amsterdam, dp 59.

Head, K. & Mayer, T. (2005), 'Market potential and the location of japanese firms in the european union', *Review of Economics and Statitistics* .

Head, K. & Mayer, T. (2006), 'Regional wage and employment responses to market potential in the eu', *Regional Science and Urban Economics* .

Head, K. & Ries, J. (2001), 'Overseas investments and firms exports', *Review of International Economics* **9**(1), 108–122.

Heiss, F. (2002), 'Structural choice analysis with nested logit', *Stata Journal* **2**(3), 227–252.

Hivatal, K. S. (2001), Ágazati kapcsolatok mérlege 1998 (input-output tables), Technical report, Központi Statisztikai Hivatal.

Holl, A. (2004), 'Manufacturing location and impacts of road transport infrastructure:empirical evidence from spain', *Regional Science and Urban Economics* **34**(3), 341–363.

Hummels, D. (2000), Toward a geography of trade costs. Univesrity of Chicago.

Javorcik, B. S. (2004), 'Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages', *American Economic Review* **94**(3), 605–627.

Katay, G. & Wolf, Z. (2004), Investment behavior, user cost and monetary policy transmission - the case of hungary, Working Paper 12, MNB.

Katay, G. & Wolf, Z. (2006), Estimating aggregate productivity from firm-level data: the case of hungary. Magyar Nemzeti Bank.

Kim, S.-H., Pickton, T. S. & Gerking, S. (2003), 'Foreign direct investment: Agglomeration economies and returns to promotion expenditures', *Review of Regional Studies* **33**(1), 61–72.

Koenker, R. & Bassett, G. (1978), 'Regression quantiles', *Econometrica* **50**, 43–61.

Koenker, R. & Hallock, K. (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**(4), 143–156.

Konings, J. (2001), 'The effects of foreign direct investment on domestic firms: Evidence from firm level panel data in emerging economies', *Economics of Transition* **9**(3).

Kosova, R. (2006), Do foreign firms crowd out domestic firms? evidence from the czech republic. School of Business and Public Management - George Washington University.

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

Luger, M. I. (1987), The states and industrial development program: Program mix and policy effectiveness, *in* J. Quigley, ed., 'Perspective on Local Public Finance and Public Policy', Jai Press, Greenwich, Conn, pp. 29–63.

Maddala, G. S. (1983), *Limited Dependent and Qualitiative Variables in Econometrics*, Cambridge University 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.

Melitz, M. J. (2004), 'The impact of trade in intra-industry reallocations and aggregate industry productivity', *Econometrica* .

Németh, N. (2004), A kistérségi tagoltság regresszióelemzése (regression analysis of regional differences). MTA KTI.

Nickell, S. (1996), 'Competition and corporate performance', *Journal of Political Economy* **104**, 724–746.

Olley, G. & Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**, 1263–1297.

Ottaviano, G., Tabuchi, T. & Thisse, J.-F. (2002), 'Agglomeration and trade revisited', *International Economic Review* **43**, 409–436.

Papke, L. E. (1989), Interstate business tax differentials and new firm location, WP 3184, NBER.

Puga, D. (2002), 'European regional policies in light of recent location theories', *Journal of Economic Geography* **2**(4), 372–406.

Redding, S. & Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of International Economics* **62**(1), 53–82.

Sabirianova, K., Svejnar, J. & Terrell, K. (2005), 'Distance to the efficiency frontier and fdi spillovers', *Journal of the European Economic Association, Papers and Proceedings* **3**(2-3), 576–586.

Shapiro, S. & Francia, R. (1972), 'An approximate analysis of variance test for normality', *Journal of the American Statistical Association* **67**, 215–216.

Sjöholm, F. (1999), 'Technology gap, competition and spillovers from direct foreign investment', *Journal of Development Studies* **36**(1), 53–73.

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

Telegdy, Á. (2004), Apeh dataset cleaning steps, Labor project, Central European University.

Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge, MA.

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**, 100–101.

White, H. (1980), 'A heteroscedasticity-consistent covariance matrix estimator and: a direct test for heteroscedasticity', *Econometrica* **48**, 817–830.

Woodward, D. P. (1992), 'Locational determinants of jqpanese manufacturing start-ups in the united states', *The Southern Economic Journal* **58**(3), 690–708.

Wooldridge, J. (2002*a*), *Econometric Analysis of Cross-Section and Panel Data*, MIT Press, Cambridge.

Wooldridge, J. (2003), 'Cluster-sample methods in applied econometrics', *American Economic Review* **93**, 133–138.

Wooldridge, J. M. (2002*b*), *Econometric Analysis of Cross Section and Panel data*, MIT Press.

Zahra, S. & George, G. (2002), 'Absorptive capacity: A review, reconceptualization, and extension', *Academy of Management Review* **27**, 185–203.