European market integration and the determinants of firm localization – the case of Poland

Agnieszka Gehringer*♦

Astrid Krenz*

Abstract

The paper analyses empirically the determinants of firms' localization in Poland. We use regional data of the sixteen Polish administrative regions over the period 2003 to 2010 to examine which role agglomeration forces and other factors played in explaining the choice to operate in a certain location. Our results suggest that agglomeration economies stemming in particular from the R&D sector, as well as human capital and the infrastructure positively influence the regional localization of firms. Poland's accession to the European Union had a positive impact for the location decision of new firms in the Polish economy.

Keywords: Localization, Agglomeration Economies, Knowledge Externalities, Polish Regions, European Integration

JEL classification: F14, F15, F23, R11, R12

* University of Goettingen, Department of Economics
♦ corresponding author: Platz der Goettinger Sieben 3
D-37073 Goettingen
Germany
Tel. 0049-(0)-551 39 33932
Fax. 0049-(0)-551 39 7093

1. Introduction

Since the end of the 19th century, the analysis of firm localization and of factors that contribute to a certain pattern of geographic concentration of economic activity has attracted attention not only of scientists, but also of policy-makers. Starting with the early contributions of location theory, paving the way to urban and regional economics, culminating in the influential contribution about externalities by Marshall (1890) and more recently with the lively development in the field of New Economic Geography (NEG), authors were interested in investigating the determinants and consequences of spatial concentration. This bulk of interest is mainly due to important implications in terms of internal and international competitiveness of industries as well as in terms of distribution of income. Moreover, the establishment of a common European market, with the aim to eliminate internal barriers of different nature, implies an environment where the standard localization determinants operate differently than usual or where new localization forces assume importance.

The past empirical investigations of industrial location within Europe are rather scarce and dedicated especially to the old member states: Figueiredo *et al.* (2002) analyze the localization determinants for Portugal, Arauzo-Carod and Manjón-Antolín (2004) for Spain, Autant-Bernard (2006) for France. For Germany, an investigation by Bade and Nerlinger (2000) concentrates on the localization of firms from new-tech based sectors in the time period 1989 to 1996. Regarding the new EU member states from Eastern Europe, the only attempts to measure localization forces were provided by Cieślik (2005a and 2005b) who investigates the determinants of localization of the foreign multinational enterprises (MNEs) in Poland. Similarly to the two papers by Cieślik, Gauselman and Marek (2011) compare the factors determining the location choice of the MNEs in 33 regions of East Germany, Poland and the Czech Republic. Consequently, barely did past investigations try to find the general determinants of industrial location in the new member states.

Our contribution is aimed at filling this gap, by analyzing the dynamics of localization of all firms in Poland in the last decade. The choice of the country is not casual. Poland carries an exemplar experience of a formerly planned economy that successfully managed the passage towards a market economy within a relatively short period of time. The process of transition involved the establishment of a new economic and institutional framework, with a dynamically progressing privatization and

industrialization and the intense restructuring of enterprises (Carlin *et al*., 1995; Fidrmuc, 2007; Robinson, 2004). These intensive efforts to re-establish a market-based economic structure - which in many areas of transformation were made shock-therapeutically rather than gradually, as in many other transition economies - were crucial in fulfilling the economic and legal requirements deriving from the EU-accession rules. During the entire transformation period since 1989 Poland became the main FDI receiving country in the region of Eastern and Central Europe (Cieślik, 2005b). Finally, in 2004 Poland has become a formal member of the EU. Based on such developments, Poland is an interesting case to investigate which factors played a crucial role in shaping the localization space of the economic activities across sectors over time.

The main contribution of our paper consists in focusing on the recent regional experience of one of the new EU member states, Poland, being the biggest among the newly accessed Eastern and Southern European members. Whereas the main focus of the past contributions was on the decisions to locate taken by the foreign investors (MNEs), we concentrate on the domestic activity as such, without making the distinction between the activity performed by the national or foreign firms. This focus is particularly important from the point of view of the European cohesion policy goals, which aim to improve certain regional indicators, such as research and development performance, regional competitiveness and business environment, the share of tertiary educational attainment, in addition to some other sustainability and inclusiveness goals. We specifically investigate the importance of R&D intensity and innovation activity for firms' localization, examine the effects due to Poland's accession to the EU and differentiate the effects for all and only the new firms in the economy. Our aim is to assess whether the determinants of localization in Poland differ from those observed for the old EU members, extensively investigated in the past.

The paper is organized as follows. In the next section, we theoretically motivate our empirical work by reviewing the past contributions. Section 3 discusses the data and the empirical methodology used. Section 4 presents and comments on the results. Finally, the last section concludes.

2. The role of agglomeration economies and other factors for the geographic localization of firms

One of the earliest contributions to draw attention on the dynamics of geographic concentration between regions was by Alfred Marshall (1890). His conceptual contribution brought to determine a

separate category of external economies of scale, consisting in (at least partially) unpaid advantages to the firms located in a certain region. Such advantages, more precisely, derive from the presence of inter-linkages in the business activities with the other firms located in the same region. Thus, such economies of scale are not internal, but external to a given firm. Subsequently, the importance of economic linkages between firms at the regional level has been extensively analyzed in the field of the NEG pioneered by Paul Krugman (1991).

More precisely, due to the proximity of upstream and downstream firms as well as of workers, each single firm experiences a unit cost advantage in an analogous manner as in the case of internal economies of scale. There are at least three different sources of such unit cost reduction. First, from the supply side, being close to suppliers will reduce input factor costs for producers who save transportation and transaction costs. Analogously, from the demand side, being close to demand will induce a reduction in transport costs as well as further benefits of improved market access for producers. Second, the proximity between the suppliers of the production factors (capital and labor) and their users makes it possible for the former to develop specialized technical properties and skills necessary for a more efficient production of the latter. Third, being located together leads to a more intensive exploitation of knowledge externalities, generated both at the demand- and at the supply-side (Gehringer 2013).

Moreover, localization benefits might refer to institutional advantages, related, for instance, to the quality of infrastructure. According to Krugman (1996), however, such positive effects are only conditionally available. Indeed, apart from centripetal forces, attracting firms to locate in a certain region, there could be opposite, centrifugal forces, that would drive the firms out of the local centre. Among the centrifugal factors, the level and intensity of congestion or the relatively high unit input prices might be claimed to exercise a negative influence on the geographic concentration.

Subsequent theoretical work in this area focused on the more precise causes of agglomeration. Helsley and Strange (1990) demonstrated that agglomeration economies can arise from beneficial pooling of specialized labor market forces. The contribution of Goldstein and Gronberg (1984) focuses on functionality advantages deriving from the opportunity to share the same supplier. Finally, Glaeser

(1999) assigns a particular role played by localized knowledge spillovers in promoting agglomeration dynamics.[1]

Against the well-developed theoretical basis underlying the special localization of economic activities, the empirical investigations of the location issue are signed by much shorter history. After the contribution of Marshall, only scarce attention has been paid to verifying the hypothesis, with the representative attempts by Carlton (1983) and Luger and Shetty (1985). Only after the significant improvement made with the emergence of the NEG, new motivation and innovative empirical approaches enriched the bulk of evidence confirming the positive role played by agglomeration economies (Head *et al*., 1995; Head and Ries, 1996; Broadman and Sun, 1997; Guimarães *et al*., 2000; Cieślik, 2005a). [2]

It is clear thus that agglomeration economies refer to different regional characteristics. This notwithstanding, there have been some attempts to use a single all-compressing measure, expressing the region's volume of economic activity. This approach encountered opposition of authors arguing that by averaging out different aspects of regional economic activity the measurement of agglomeration dynamics becomes imprecise (Head *et al*., 1995; Guimarães *et al*., 2000). In a recent investigation, Cieślik (2005a) faces such concerns by distinguishing between four types of agglomeration economies: *overall agglomeration economies*, approximated by the regional GDP volume and measuring both the demand- and supply-side regional economic volume (Head and Ries, 1996; Broadman and Sun, 1997); *urbanization economies* measured by the percentage share of the population living in the urban areas and expressing the potential informational advantage of regions with high urban density (Glickman and Woodward, 1988; Coughlin and Segev, 2000)[3]; *industry* and *service specific economies*, respectively, expressed by the respective shares of employment in the secondary/tertiary sector and representing the specialization advantage in a specific industrial or service sector (Smith and Florida, 1994; Guimarães *et al*., 2000; Woodward, 2000).

When investigating the localization forces, we include the standard determinants as well as some new factors, relating particularly to the generation and diffusion of knowledge. Among the standard

[1] A comprehensive survey on micro-founded analyses of agglomeration economies is offered by Quigley (1998).
[2] See Arauzo-Carod *et al.* (2010) for a summary of recent empirical contributions.
[3] Cieślik (2005a) observes, however, that this type of agglomeration externalities might have a centrifugal content, considering that the strong urbanization dynamics leads to intensified congestion and pollution pressure.

determinants, agglomeration forces –as described above-- belong to the group of factors that are prevalently taken into consideration (Bellak *et al*., 2008; Barrios *et al*., 2006). Additionally, other forces related to the labor market and to some other regional characteristics have been intensively investigated, especially in the context of the localization decision connected with the FDI inflows (Chung and Alcácer, 2002; Guimarães *et al*. 2000; Cieślik, 2005a; Barrios *et al*., 2006; Chidlow *et al*., 2009; Hilber and Voicu, 2010). Whereas agglomeration economies have often been confirmed to play a significant role in positively influencing the regional attractiveness, the evidence regarding the labor market forces - at least for the industrialized countries - was rather mixed. This regards more specifically the characteristics of the labor force, its availability and the educational attainment (Hilber and Voicu, 2010). The past estimations confirmed that the higher the level of unit wages the lower the attractiveness of the region for the location purposes (Crozet *et al.*, 2004). Regarding the level of education, the literature demonstrates positive effects of human capital, as proxied by the level of educational attainment among the local labor force (Alama-Sabater et al., 2011; Arauzo-Carod, 2013; Cheng and Stough, 2006; Coughlin and Sergev 2000; Coughlin et al. 1991; Egeln et al. 2004), although a negative relationship has been found as well (Arauzo-Carod, 2005; Arauzo-Carod and Viladecans-Marsal, 2009). Arauzo-Carod (2013) finds evidence for a positive relation between the level of educational attainment and firms' localization decision when measuring human capital over a wider spatial area. Arguably, employers look out for potential employees not only in the local area but also in the surroundings.

Finally, the distinction between newly establishing and incumbent enterprises seems to be non negligible. Indeed, the literature found a different pattern of localization decisions between the newly established and the already existing firms. In particular, the Marshallian forces were found to be less relevant for the location decision of the new firms (Rosenthal and Strange 2001). Different effects between start-up and relocating firms are also found by Holl (2004) an Arauzo Carod and Manjón Antólin (2011). Specifically, the study of Arouzo Carod and Manjón Antólin finds that factors of external economies matter more for the start-up than for the relocating firms.

3. Empirical Set-up

3.1 Data Issues and Selection of Variables

For our analysis we mainly use data from the Statistical Regional Yearbooks of the Polish Central Statistical Office's online database. NUTS II data were retrieved, which gave us observations for the 16 Polish voivodeships for the years from 2002 to 2010. From the Eurostat's online database we extracted data on the number of patents per labor workforce.[4]

Our dependent variable is a count measuring the number of firms (entities of the national economy recorded in the firms' REGON register) operating in a given region. These firms are enrolled in the Polish commercial register, the *Krajowy Rejestr Sądowy*.

To explain the localization behavior of firms across different regions of the Polish economy, in a first step we follow the past literature and apply a set of standard determinants. These refer to various measures at the regional level: GDP, industry agglomeration, services agglomeration, urbanization, wage level, unemployment, human capital, land area, and the quality of the infrastructure (see, for example, Cieślik, 2005a and 2013). We additionally control for R&D agglomeration economies and the patents per workforce. The following paragraphs outline the precise choice and construction of the variables.

The value of the region-level GDP is intended to capture the economic size of market demand. A higher level of regional GDP, leading consequently to higher consumer expenditures in that region, can be expected to provide an incentive for the firms to choose that location.[5]

As illustrated in the previous section, different agglomeration factors could play a relevant role in explaining firms' location choices. First, we measure agglomeration economies separately for industrial and service sectors. They are expressed as the respective shares of employment in industry and services over the total employment in the region. The stronger the presence of industrial and/or service sectors in a region, the better represented are the potential suppliers and, consequently, the higher the expected benefits from locating in that region.

[4] The full list of variables with their description is contained in Table A.1 of the Appendix. In Table A.2 of the Appendix, we show basic summary statistics.
[5] All variables that are denoted in values were converted into constant 2002 prices using the regional consumer price index for goods and services.

Second, the quality of the infrastructure might also positively influence the location choice. More precisely, the improvement of infrastructure increases the economic potential of regions and enhances agglomeration forces (Cieślik and Rokicki, 2013). Moreover, a high quality of the infrastructure reduces transportation time and distance, facilitating firms in reaching both the suppliers and the consumers. In accordance with the literature, the quality of the infrastructure is measured in terms of the density of the road network, railway lines network, share of telephone lines per population and a region's area. We can expect a positive influence for the road network, railway lines and telephone lines on the firms' localization, however, the influence of a region's area is not clear. On the one hand, a larger area will offer more space to localize production plants. On the other hand, a larger area would imply that the transportation network density will be smaller (Cieślik, 2005a).

Third, the past literature has recognized that urbanization economies might play an important role in determining localization, but the direction of influence is still a matter of discussion (Cieślik, 2005a). On the one hand, urbanization might yield benefits via labor market pooling, an improved infrastructure and means of better information networks. On the other hand, however, dense localization might also generate congestion costs, deriving, for example, from pollution, traffic jams or social distress. Such negative factors might, consequently, discourage the choice to locate in highly urbanized regions. We measure urbanization effects in a standard way, as a percentage share of urban population over overall (urban and rural) population in a region.

As a novelty to the agglomeration literature for the new member states, we implement new agglomeration factors, referring more precisely to the local innovative efforts and locally available knowledge base. In particular, we introduce a variable measuring agglomeration economies stemming from the R&D sector. This variable is constructed as the share of the R&D sector's employment over the total region's employment. The R&D sector is considered here as a measure of the region's overall effort to enhance the local degree of innovativeness (Mudambi and Swift 2012). Given, however, the technological and commercial uncertainty to actually transform such innovative input into an innovative and marketable output, R&D activities do not necessarily reflect the readily available knowledge base. This notwithstanding, firms observing a relatively high R&D intensity in a region should be attracted to locate. Instead to approximate the actual innovative output of a region, we

include also the number of patents relative to the workforce: the higher the number of patents in a region the more attractive for the firms it is to locate and to take advantage from potential knowledge spillovers.[6]

Further variables are designed to capture labor market effects. In particular, a higher wage level induces higher labor costs for potential employers and will, consequently, reduce their incentive to set up business activity in a region. At the same time, higher wages might act as an indicator for the quality of the labor force, and the more likely will firms localize in the respective region (Cieślik 2005a). Thus, the net effect of the wage level on localization is theoretically unclear. The measure we take is average monthly gross wages and salaries.

Another labor market factor refers to the regional unemployment rate. A high unemployment rate (relative to the other regions) might indicate a region's economic decline, and as such will be detrimental to firms' localization decisions. Our measure here is the average annual unemployment rate for the overall active population.

Finally, the share of students per population serves as a measure of a region's educational level. A higher share of students serves as an indicator for a region's ability to increase the potential workforce qualification, and degree of knowledge spillovers. A firm deciding where to locate will judge positively on this regional attribute and will be more likely to localize in that region. With our indicator, we follow the literature, measuring the stock of human capital in terms of labor force that completed secondary and/or tertiary level of education (Alama-Sabater, *et al*. 2011; Arauzo-Carod, *et al*., 2010; Coughlin and Sergev, 2000; Coughlin *et al*., 1991).

Furthermore, to capture the influence of the unobservable factors referring to regional characteristics, we introduce regional dummy variables (for the regions in the North-West, North, South-West, South, East; and West is the reference category), as well as time dummies to account for time-specific effects. Alternatively to the regional dummies, we introduce a measure capturing the distance to the German border, measured as the distance of a Polish region's capital city on main car

[6] Also the relative number of patents as a measure of the actual knowledge base is only an approximation and, as such, an imperfect measure of the regional degree of innovativeness. Indeed, not every innovation ends up to be patented and there is a considerable stock of locally available knowledge that remains unprotected and that constitutes an even more important source of knowledge spillovers. Nevertheless, the difficulty to grasp this phenomenon by means of a single variable is non-negligible.

routes to one of the German border cities, namely, Penkun, Görlitz or Frankfurt an der Oder. Finally, we also replace the time dummies with a dummy variable capturing the time of Poland's accession to the EU, counting a zero until 2003 and a one thereafter.

3.2 Methodology

Given that our dependent variable is a count, meaning that it takes exclusively nonnegative integer values, in principle we would need to employ count-data-regression methods. In econometric theory, count data are mostly modeled by a Poisson distribution. Other possible methods comprise the binomial or negative binomial distribution. Whereas the benefit of modeling a binomial distribution is the capability to deal with an upper bound of counts, the negative binomial distribution can ease conditional moments restrictions imposed by the Poisson distribution, as we will see in the following.

Under the assumption of a Poisson distribution for a given count variable y and a given set of explanatory variables X, the density can be written as (Wooldridge, 2002):

$$f(y|X) = \exp[-\mu(X)]\,[\mu(X)]^{y}/y! \quad (1)$$

where $\mu(X) \equiv E(y|X)$ denotes the conditional mean. A conditional maximum likelihood estimator can be derived. This estimator will be efficient. A strong restriction, however, is imposed on the conditional moments, namely the equality between the conditional variance and mean.

In our context, the number of firms y_i is drawn from a Poisson distribution with parameter $\mu\,(x_i)$ and it is dependent on a set of regional variables x_i, contained in the vector X_i,such that the probability to observe a count of firms is:

$$P(y_i|X_i) = \exp[-\mu(X_i)]\,[\mu(X_i)]^{y_i}/y_i \quad (2)$$

The most common functional form taken for the mean is the exponential function, such that $\mu(X_i) = \exp(X_i\beta)$. Vector β is the column vector of coefficients on regional explanatory variables which has to be estimated.

To circumvent the problematic moment restrictions for the Poisson models, econometric theory offers for example the NegBin II model of Cameron and Trivedi (1986), which is a negative binomial regression model. In this model, an additional term c_i capturing unobserved heterogeneity is considered. It is assumed that c_i is independent of x_i and has a gamma distribution with unit mean and

variance η^2. The conditional mean is the same as in the Poisson model, however, the variance is different (bigger than the mean) and can be written as:

$$Var(y_i|X_i) = E(y_i|X_i) + \eta^2 (E(y_i|X_i))^2 \quad (3)$$

The panel structure of our data creates a need to control for unobserved heterogeneity. In this respect, the literature has focused attention on problems arising from using both unconditional (Allison and Waterman 2002) and conditional negative binomial panel regression estimators (Greene, 2007). Allison and Waterman drew attention to the fact that the conditional fixed effects negative binomial estimator typically coded in statistical programs is not a real fixed effects model because it does not control for all stable covariates. Whereas Allison and Waterman demonstrate then within their study's simulations that the unconditional fixed effects estimator (supplementing the panel OLS estimator with fixed regional effects) works well, Greene argues that their estimator still suffers from an incidental parameters problem and thus produces inconsistent estimates. Keeping in mind this methodological issue, in our context it seems crucial to observe that the between variability of our variables dominates significantly the within variability (see Table 1 in the Appendix for the descriptive statistics). Consequently, using the fixed effect estimator, which removes the between dimension and relies exclusively on the within dimension, would detract much of the economically meaningful part of our analysis. For that reason, we opt for the pooled OLS regressions as our preferred estimation technique. In doing that, however, we add and remove – in the separate specifications – regional and time effects to control for the robustness of the results.

3.3 Descriptive Evidence

Taking a look at the spatial distribution of firms in the Polish economy in 2010, we can see that across the regions the largest number of firms (for all and the newly registered firms) is given for the regions Mazowieckie —the central area around the Polish capital city Warsaw— Slaskie, Wielkopolskie and Dolnoslaskie. Clearly, the Southern, Western and Central parts around the capital city bear the highest firm activity in the Polish economy (Table 1). These are also regions with economically important urban centers, like Katowice in Slaskie, Poznan in Wielkopolskie and Wroclaw in Dolnoslaskie. At

the same time, however, Mazowieckie remains the leader both in terms of the number of firms operating in the region and in terms of the relative volume of investment.[7]

The greatest dynamics in terms of growth of the number of firms are found for the regions Mazowieckie, Wielkopolskie and Pomorskie (with the three important harbor cities Gdansk, Gdynia and Sopot). The data reveal that there has been less growth of the number of firms in the Eastern parts of the country between 2003 and 2010. These regions continue to experience the lowest GDP per capita relative to the Polish average. In 2010, GDP per capita was still between 67% of the Polish average in Podkarpackie and Lubelskie and 73% in Podslaskie and Warminsjo-mazurkie, compared to 162% of Mazowieckie (GUS, 2013).

Table 1: Descriptive Evidence about the spatial distribution of firms in the Polish Economy

		2003		2010		Change of number of new firms between 2003 and 2010, in %
Location	Region	New firms	All firms	New firms	All firms	
South-West	Dolnośląskie	21448	305772	33257	331107	55,06
North	Kujawsko-pomorskie	13029	191096	19518	185946	49,80
East	Lubelskie	12051	154916	17647	164049	46,43
North-West	Lubuskie	7566	97348	11767	106107	55,53
Central	Łódzkie	17141	241530	25170	230853	46,84
South	Małopolskie	20494	287886	36370	331334	77,47
Central	Mazowieckie	39545	575716	66365	681032	67,82
South-West	Opolskie	5595	87409	8545	99079	52,72
East	Podkarpackie	10486	142613	15960	152574	52,20
East	Podlaskie	6749	96891	10229	91823	51,58
North	Pomorskie	17074	226334	27533	260089	61,26
South	Śląskie	29233	423877	46349	451443	58,55
East	Świętokrzyskie	6975	103715	10517	108760	50,79
North	Warmińsko-mazurskie	9145	110453	13666	119065	49,43
North-West	Wielkopolskie	22848	335657	38259	375351	67,45
North-West	Zachodniopomorskie	14247	200645	20857	220466	46,40

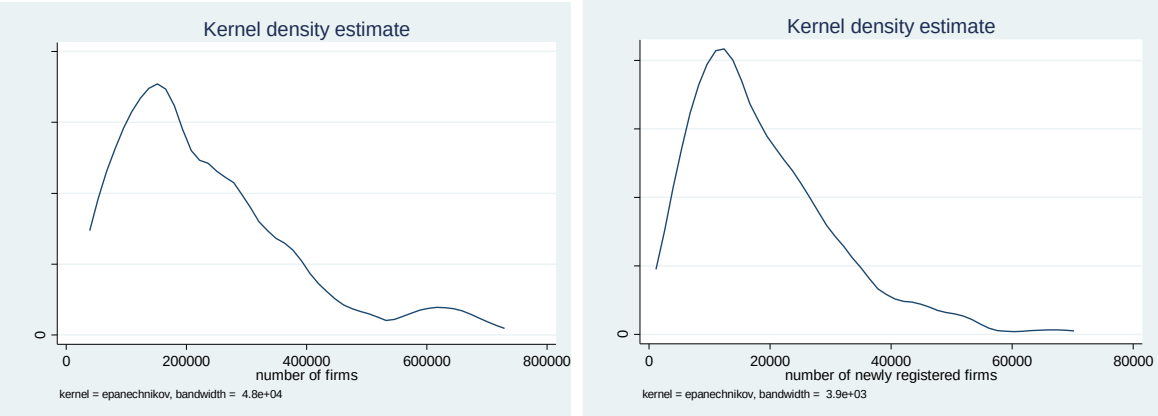
Source: Statistical Regional Yearbooks of the Polish Central Statistical Office, authors' computations.

By plotting the density of the regional counts of firms (see Figure 1), we find that for both the newly registered firms and all firms in the economy the distribution follows a left-hump-shaped line. It resembles a Poisson-process with a larger frequency of the lower count of the variable. The probability to observe a count of firms of around 180000 is highest for all registered firms for a region, whereas it

[7] In 2011, the total volume of investment of businesses in Mazowieckie amounted for 29687.6 mln zloty and constituted 22.7 % of investment made in Poland (GUS, 2013).

is less probable to find for example around 20000 or 400000 firms in a NUTS 2-region. For newly registered firms, it is most probable to observe around 18000 firms. We conclude that the distribution of number of firms (new and all together) across regions indeed can be modeled through a Poisson distribution, and because of this we are able to employ the econometric methods explained in Section 3.2.

Figure 1: Density plots for the count of firms



Source: Statistical Regional Yearbooks of the Polish Central Statistical Office, authors' computations and depictions. The plots show kernel density estimates of the number of firms across regions in the Polish economy.

4. Econometric Analyses

Our empirical analysis examines firms' localization behavior in the Polish economy from 2003 to 2010. We investigate the determinants of the localization of firms through a negative binomial regression analysis.[8] More specifically, the expected count of firms will be investigated with regard to regional explanatory characteristics. In the regression analyses, we will make use of lagged explanatory variables in order to avoid simultaneity problems. We first investigate the location determinants for all firms, those already established and those newly registered in each period. Indeed, it might be the case that for the newly registering firms some other specific factors play a role in choosing their location. Here, in particular, we are interested in the impact of the EU membership: we do expect that this factor could have motivated new firms to enter the market.

4.1 Regression results

[8] For all regression models, the negative binomial estimator proved to be preferable over the Poisson estimator, which can be seen by the alpha test results in the following regression output tables.

All registered firms

Regressions results for the sample of all firms are displayed in Table 2. The results demonstrate that industrial, services and R&D agglomeration economies significantly influence firms' localization in a given region. Note, however, that since our dependent variable measures the count share of all firms, already registered and newly registered, the interpretation of the effects should account for this. In particular, what we observe here is not only the decision to locate in a certain region, but also the ability to survive and willingness to stay in a certain location. To more precisely refer to the very decision to locate, we perform the analogous estimations like in Table 2 for the restricted sample of newly registered firms (see Table 3 below).

Other things being equal, for a one unit increase in the share of industrial employment, the expected log of count of the firms raises by 0.025 to 0.041 units over the period from 2003 to 2010. In the case of services agglomeration economies, the effect is comparable, although slightly weaker, and amounts from 0.011 up to 0.034 units. In terms of size, however, the strongest effect clearly stems from agglomeration economies in the R&D sector. Clearly, R&D activity plays an important role for the localization of firms within the Polish economy, as it crucially influences the expectation of benefits deriving from research-based linkages between firms. Similarly, the existence of high-quality human capital, as measured in terms of the number of students per population, has been confirmed as an important factor enhancing localization of businesses. At the same time, there is no evidence of the importance for the firms' location of the readily available knowledge stock, incorporated in the patented innovative output.

We further find that the quality of the road network positively and significantly influences the number of firms in a region. The road transportation is the only infrastructure that positively determines the location of firms. Instead, the effect for railways and telephone lines in the majority of cases remained insignificant. The effects arising from regional GDP and from the size of the area are also positive, but not that large. Urbanization economies show some but only weakly significant effect and not in all specifications. For wages, we could confirm the presence of a negative effect on localization.

The validity of the negative binomial regression model can be assessed by looking at the value of alpha which results from a regression testing for equidispersion (Cameron and Trivedi 2009). If alpha equals zero, the Poisson regression model is preferred over the negative binomial regression model. Our results imply that we can safely conclude that the negative binomial regression model is the adequate choice, for in every specification we reject the null hypothesis of equidispersion. Moreover, the Wald-Chi square statistic resulting from a test that all coefficients are equal to zero implies that the model specifications are all statistically significant.

Table 2: Regressions results for all registered firms

	Dependent variable: number of all registered firms in a region						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP	3.68e-06***	3.42e-06***	3.45e-06***	2.26e-06	3.36e-06***	4.08e-06***	4.00e-06***
	(1.91e-06)	(1.48e-06)	(1.47e-06)	(1.48e-06)	(1.50e-06)	(2.35e-06)	(2.72e-06)
Industry agglom. economies	0.0254***	0.0261***	0.0267***	0.0407***	0.0286***	0.0382***	0.0276***
	(0.0053)	(0.0057)	(0.0058)	(0.0103)	(0.0143)	(0.0106)	(0.0150)
Service agglom. economies	0.0136***	0.0105**	0.0111**	0.0340***	0.0123**	0.0402***	0.0122**
	(0.0080)	(0.0088)	(0.0088)	(0.0062)	(0.0013)	(0.0092)	(0.0133)
R&D agglom. economies		0.6230***	0.585***	0.907***	0.572***	0.945***	0.627***
		(0.316)	(0.317)	(0.355)	(0.291)	(0.331)	(0.171)
Urbanization economies	0.0047	0.0071**	0.0066*	0.0059*	0.0065*	0.0024	0.0056
	(0.0074)	(0.0077)	(0.0079)	(0.0060)	(0.0077)	(0.0071)	(0.0089)
Wages	-0.0002	-0.0004***	-0.0004***	-0.0007***	-0.0004***	-0.0011***	-0.0005**
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
Unemployment rate	0.0012	0.0039	0.0019	-0.0156**	0.0023	-0.0302***	0.0024
	(0.0064)	(0.0040)	(0.0043)	(0.0062)	(0.0053)	(0.0011)	(0.007485)
Students per population	0.111***	0.0841***	0.0923***	0.122***	0.0947***	0.169***	0.107***
	(0.0423)	(0.0035)	(0.0037)	(0.0051)	(0.0035)	(0.0059)	(0.0036)
Patents per workforce		-0.0005	-0.0006	0.0005	-0.0003	-0.0007	-0.0008
		(0.0045)	(0.0043)	(0.0025)	(0.0036)	(0.0024)	(0.0032)
Telephone lines	0.0002	-0.0143**	-0.0132**	-0.0042	-0.0131**	0.0042	-0.0123
	(0.0110)	(0.0079)	(0.0086)	(0.0093)	(0.0083)	(0.0012)	(0.0013)
Road network	0.0084***	0.0071***	0.0069***	0.0130***	0.0069***	0.0118***	0.0059***
	(0.0028)	(0.0032)	(0.0032)	(0.0045)	(0.0031)	(0.0054)	(0.0032)
Railway lines	-0.0125	-0.0062	-0.0057	-0.0021	-0.0065	0.0167	0.0012
	(0.0255)	(0.0274)	(0.0274)	(0.0296)	(0.0273)	(0.0390)	(0.0294)
Area	3.78e-05***	3.50e-05***	3.45e-05***	3.67e-05***	3.50e-05***	3.01e-05***	3.28e-05***
	(9.96e-06)	(8.77e-06)	(9.07e-06)	(9.23e-06)	(1.00e-05)	(1.09e-05)	(1.04e-05)
EU membership			-0.0375	-0.0733**	-0.0379		
			(0.0291)	(0.0373)	(0.0289)		
GE-PL Border					7.46e-05		8.63e-05
					(0.0005)		(0.0005)
Constant	8.839***	9.554***	9.552***	7.873***	9.398***	8.482***	9.685***
	(0.649)	(0.539)	(0.521)	(0.749)	(0.936)	(1.100)	(1.204)
Reg. dummies	no	no	no	yes	no	yes	no
Time dummies	no	no	no	no	no	yes	yes
Log likelihood	-1501.3	-1443.4	-1442.9	-1417.39	-1442.7	-1410.9	-1441.9
Alpha	0.0235	0.0191	0.0190	0.0126	0.0189	0.0114	0.0187
(s.e.)	(0.0030)	(0.0025)	(0.0024)	(0.0015)	(0.0025)	(0.0013)	(0.0025)
Wald Chi-sq.	3304.7	4244.9	4190.3	5992.0	4756.7	6901.9	4846.5
[p-value]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Nr. Obs.	128	124	124	124	124	124	124

Source: Authors' computations based on the Statistical Regional Yearbooks of the Polish Central Statistical Office and Eurostat.
Notes: The results refer to negative binomial estimates on the pooled data. Cluster-robust standard errors are displayed in parentheses. The table reports the log likelihood, the value alpha resulting from a regression testing for equidispersion , which gives

a value of zero if the Poisson regression model is preferred over the negative binomial regression model, the respective robust standard errors, the Wald Chi-square statistic and the respective p-value. * denotes significance at a 10% level, ** denotes significance at a 5% level, *** denotes significance at a 1% level

Newly registered firms

As we illustrated in the theoretical part, there are some reasons to expect that factors determining firms localization work differently between the newly and already established firms. Generally, we could confirm that also for the sample of newly registered firms the industry, services and R&D agglomeration economies as well as the share of students per population and the road network are important explanatory factors (Table 3). There are, however, some peculiarities compared to the case of all registered firms. In particular, the size of the market, as approximated by regional GDP becomes insignificant, but in two last specifications. This suggests that a higher market demand is rather more important for the regional localization choice of incumbent firms than of the new firms which just enter the market. It seems thus that for the already established firms it is crucial for their survival to meet a sufficient market size. For the entering firms, it can well be that the market is still under formation, so that its initial size is not significantly determining their decision to enter. Moreover, effects due to urbanization and the EU membership in most specifications become positively significant. For the EU membership, the contrast with the estimations regarding all firms is comprehensible. When estimating the regressions for newly registered firms, we are grasping the effect of firms actually deciding to localize, so that the opportunities to exploit the enlarged common market could have motivated more new establishments in the Polish regions. The EU membership has thus increased the attractiveness of the Polish regions for the new investors. Finally, we find a positive influence on the location decision of newly registering firms due to the proximity to the German border. Again, from the results of the alpha-test we conclude that negative binomial estimators are preferred over Poisson estimators. Moreover, the model specifications appear to be statistically significant.

Table 3: Regression results for the newly registered firms

	Dependent variable: number of all registered firms in a region						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP	1.67e-06	1.32e-06	1.28e-06	1.25e-06	8.89e-07	4.58e-06***	3.56e-06**
	(1.82e-06)	(1.32e-06)	(1.34e-06)	(1.17e-06)	(1.27e-06)	(2.23e-06)	(2.75e-06)
Industry agglom. economies	0.0299***	0.0288***	0.0274***	0.0533***	0.0370***	0.0452***	0.0321***
	(0.0063)	(0.0065)	(0.0065)	(0.0122)	(0.0145)	(0.0107)	(0.0152)
Service agglom. economies	0.0202***	0.0163***	0.0151***	0.0334***	0.0206***	0.0409***	0.0214**
	(0.0075)	(0.0076)	(0.0075)	(0.0066)	(0.0134)	(0.0082)	(0.0125)
R&D agglom. economies		0.606***	0.648***	0.909***	0.623***	1.021***	0.830***
		(0.395)	(0.403)	(0.401)	(0.343)	(0.302)	(0.193)
Urbanization economies	0.0101**	0.0125***	0.0137***	0.0138***	0.0135***	0.0021	0.0070
	(0.0104)	(0.0104)	(0.0103)	(0.0072)	(0.0093)	(0.0072)	(0.0096)
Wages	-9.75e-05	-0.0002	-0.0002*	-0.0005***	-0.0003*	-0.0014***	-0.0009***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
Unemployment rate	-0.0058	0.0023	0.0017	-0.0089	0.0035	-0.0258***	0.0048
	(0.0075)	(0.0058)	(0.0061)	(0.0073)	(0.0069)	(0.0082)	(0.0099)
Students per population	0.122***	0.0901***	0.0732**	0.121**	0.0843***	0.220***	0.127***
	(0.048)	(0.0512)	(0.0504)	(0.055)	(0.0445)	(0.062)	(0.043)
Patents per workforce		0.0029	0.0033	0.0055**	0.0047	0.0022	0.0022
		(0.0038)	(0.0038)	(0.0026)	(0.0032)	(0.0021)	(0.0031)
Telephone lines	-0.0174**	-0.0314***	-0.0336***	-0.0254***	-0.0329***	0.0046	-0.0172**
	(0.0121)	(0.0085)	(0.0089)	(0.0114)	(0.0093)	(0.0108)	(0.0124)
Road network	0.0096***	0.0079***	0.0083***	0.0107***	0.0084***	0.0072**	0.0044***
	(0.0038)	(0.0043)	(0.0043)	(0.0043)	(0.0039)	(0.0048)	(0.0034)
Railway lines	-0.0281*	-0.0017	-0.0018	-0.0030	-0.0022	0.0269	0.0131
	(0.0374)	(0.0379)	(0.0378)	(0.0345)	(0.0366)	(0.0392)	(0.0336)
Area	4.58e-05***	4.36e-05***	4.46e-05***	4.44e-05***	4.70e-05***	3.17e-05***	3.84e-05***
	(1.08e-05)	(1.15e-05)	(1.12e-05)	(8.86e-06)	(1.13e-05)	(8.86e-06)	(9.92e-06)
EU membership			0.0779*	0.0363	0.0751*		
			(0.0259)	(0.0417)	(0.0286)		
GE-PL Border					0.0004*		0.0004**
					(0.0005)		(0.0005)
Constant	5.725***	6.493***	6.506***	4.761***	5.771***	5.946***	6.841***
	(0.688)	(0.640)	(0.661)	(0.803)	(1.116)	(1.011)	(1.255)
Reg. dummies	no	no	no	yes	no	yes	no
Time dummies	no	no	no	no	no	yes	yes
Log likelihood	-1196.1	-1150.2	-1148.6	-1124.6	-1145.5	-1102.2	-1138.6
Alpha	0.0302	0.0254	0.0248	0.0168	0.0236	0.0117	0.0211
(s.e.)	(0.0039)	(0.0032)	(0.0031)	(0.0021)	(0.0029)	(0.0014)	(0.0028)
Wald Chi-sq.	2321.8	2532.8	2603.8	3270.1	2952.2	6485.1	3675.3
[p-value]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Nr. Obs.	128	124	124	124	124	124	124

Source: Authors' computations based on the Statistical Regional Yearbooks of the Polish Central Statistical Office and Eurostat.
Notes: The results refer to negative binomial estimates on the pooled data. Cluster-robust standard errors are displayed in parentheses. The table reports the log likelihood, the value alpha resulting from a regression testing for equidispersion which gives a value of zero if the Poisson regression model is preferred over the negative binomial regression model, the respective robust standard errors, the Wald Chi-square statistic and the respective p-value. * denotes significance at a 10% level, ** denotes significance at a 5% level, *** denotes significance at a 1% level.

4.2 Further Robustness Checks

In the former section, we already established the robustness of our results based on the inclusion of different factors accounting for the regional and time fixed effects. In this section, we additionally check for the robustness of the results by comparing the negative binomial estimators with estimates

from a basic pooled OLS regression. As discussed before, the negative binomial method has the advantage of accounting for the nature of the dependent variable that is a count – something that the simple OLS methodology does not. By running OLS estimation we want to check which differences in the results can be observed when assuming an unconditional linear relation between our dependent and the explanatory variables.

Results in Table 4 confirm the relationships already discussed above: we find highly significant positive effects due to agglomeration economies and the share of students, no matter if we additionally control for regional or time effects or if we just exclude them. Additionally, wages are found to exert a significantly negative influence on firms' localization in most of the specifications. As regards the effects for GDP, the road network, the EU accession and the German-Polish border, OLS regression results generally reveal no significant effects.[9] The effects are significant, when employing adequate estimation methodologies: the count data regression framework that, in our case, is given by the negative binomial estimators.

Table 4: Robustness Analysis

	Ln(newfirms)			Ln(allfirms)		
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	3.87e-06	4.81e-06*	1.32e-06	4.18e-06	4.40e-06	2.46e-06
	(3.31e-06)	(2.58e-06)	(1.32e-06)	(3.22e-06)	(2.68e-06)	(1.65e-06)
Industry agglom. economies	0.0296	0.0440***	0.0526***	0.0251	0.0367***	0.0393***
	(0.0177)	(0.0126)	(0.0139)	(0.0170)	(0.0124)	(0.0120)
Services agglom. economies	0.0198	0.0406***	0.0333***	0.0099	0.0399***	0.0340***
	(0.0132)	(0.0093)	(0.0076)	(0.0142)	(0.0105)	(0.0073)
R&D agglom. economies	0.809***	1.020***	0.915**	0.626***	0.961**	0.914**
	(0.234)	(0.330)	(0.416)	(0.207)	(0.358)	(0.368)
Urbanization economies	0.0075	0.0017	0.0135*	0.0059	0.0021	0.0061
	(0.0106)	(0.0079)	(0.0075)	(0.0102)	(0.0078)	(0.0064)
Wages	-0.0009	-0.0014***	-0.00054**	-0.0005	-0.0012**	-0.0007***
	(0.0006)	(0.0004)	(0.0002)	(0.0005)	(0.0004)	(0.0002)
Unemployment rate	0.0059	-0.0256**	-0.0094	0.0027	-0.0302**	-0.0165**
	(0.0106)	(0.0093)	(0.0076)	(0.0092)	(0.0117)	(0.0065)
Students per population	0.123**	0.220***	0.124*	0.106**	0.171**	0.122**
	(0.0484)	(0.0687)	(0.0608)	(0.0393)	(0.0641)	(0.0557)
Patents per workforce	0.0026	0.0022	0.0058**	-0.0004	-0.0006	0.0009
	(0.0033)	(0.0023)	(0.0027)	(0.0033)	(0.0028)	(0.0027)
Telephone lines	-0.0150	0.0056	-0.0254**	-0.0109	0.0049	-0.0045
	(0.0145)	(0.0122)	(0.0119)	(0.0140)	(0.0135)	(0.0102)
Road network	0.0048	0.0071	0.0106**	0.0061	0.0115*	0.0131**
	(0.0038)	(0.0052)	(0.0043)	(0.0037)	(0.0058)	(0.0047)
Railway lines	0.0120	0.0288	-0.0276	0.0024	0.0200	-0.0004
	(0.0362)	(0.0425)	(0.0373)	(0.0328)	(0.0418)	(0.0316)
Area	3.86e-05***	3.13e-05***	4.40e-05***	3.25e-05***	2.92e-05**	3.63e-05***
	(1.01e-05)	(9.52e-06)	(9.37e-06)	(1.08e-05)	(1.16e-05)	(9.57e-06)
GE-PL border	0.0004			1.89e-05		
	(0.0005)			(0.000475)		
EU membership			0.0391			-0.0709*

[9] We have also run panel conditional fixed effects negative binomial estimators. However, we faced convergence problems, a fact that used to appear in other studies as well (see Arauzo Carod and Manjón Antólin (2011)). Given that most of our attention lies on the between variation of the data, we work with the pooled panel OLS estimates in a same way as has been done for example by Cieslik (2005a).

			(0.0435)			(0.0392)
Constant	6.956***	6.039***	4.813***	9.827***	8.621***	7.966***
	(1.413)	(1.133)	(0.815)	(1.318)	(1.212)	(0.772)
Regional dummies	no	yes	yes	no	yes	yes
Time dummies	yes	yes	no	yes	yes	no
R^2	0.938	0.966	0.951	0.942	0.965	0.961
Nr. obs.	124	124	124	124	124	124

Source: Statistical Regional Yearbooks of the Polish Central Statistical Office and Eurostat, authors' computations.
Notes: We show results from running pooled OLS regressions. Cluster-robust standard errors are displayed in parentheses. * denotes significance at a 10% level, ** denotes significance at a 5% level, *** denotes significance at a 1% level.

5. Conclusions

In this paper we investigated the determinants of the localization of firms in the Polish economy. We extended the existing literature by a) controlling for additional variables capturing the effects of R&D intensity and innovation activity for the localization behavior of firms, b) taking into account the effects due to Poland's accession into the EU, and c) differentiating the effects for all and only the newly establishing firms in the economy.

We found that firms' localization is positively influenced by industry and services agglomeration economies and, most importantly, by agglomeration economies stemming from the R&D sector, by the road network and the share of students per population. In that way, we could confirm that the standard determinants of localization that have been found in the past agglomeration literature and particularly by Cieślik (2005a) in the case of Poland to play a significant role for the foreign investors are also valid for firms in general. Additionally, we demonstrate the relevance of R&D activities for localization decisions. On top of that, for the newly registered firms, positive effects due to the accession to the European Union and due to proximity to the German border could be detected.

In terms of policy implications, we focus on the two most powerful channels, namely, R&D agglomeration economies and human capital accumulation. Policy measures aimed at sustaining the proper training and skill upgrading of the local labor force, as well as at stimulating the dynamic R&D environment will offer favorable conditions for the movement of goods and production inputs and create an incentive for the localization of firm activity. Moreover, improvement of the road infrastructure should remain an important policy priority. In this respect, the European structural funds could have played a crucial role. Indeed, in the period 2007 to 2013 Poland received the largest part of financial means of the European structural funds. All firms, the domestic and multinational ones located in Poland were capable to apply for the EU funding. The use of these finances was conditioned

to be employed for the qualification of human capital, subsidies for innovative investments and the cooperation between research institutes and enterprises, among others. Given that Poland's accession to the European Union bears positive effects for the establishment of new firm activity, the transition from the planned economy to an open market economy and finally to becoming a member in the European liberal market can be seen as a dynamic process that brought about favorable EU structural support and competitive advantages.

References

Alamá-Sabater, L., Artal-Tur, A. and Navarro-Azorín, H. M. (2011). 'Industrial location, spatial discrete choice models and the need to account for neighbourhood effects', Annals of Regional Science, 47, pp. 393-418.

Alison, P. D., Waterman, R. P. (2002). 'Fixed-effects negative binomial regression models', Sociological Methodology, 32, 247-265.

Arauzo-Carod, J. M. (2013). 'Location determinants of new firms: Does skill level of human capital really matter?', Growth and Change, 44, pp. 118-148.

Arauzo-Carod, J. M. (2005). 'Determinants of industrial location: An application for Catalan municipalities', Papers in Regional Science, 84, pp. 105-120.

Arauzo-Carod, J. M. and Manjon-Antolin, M. C. (2011). 'Locations and relocations: Determinants, modeling, and interrelations', The Annals of Regional Science, 47, pp. 131-146.

Arauzo-Carod, J. M., and Manjón-Antolín, M. C. (2004). 'Firm size and geographical aggregation: An empirical appraisal in industrial location', Small Business Economics, 22, pp. 299-312.

Arauzo-Carod, J. M., Liviano-Solis, D. and Manjon-Antolin, M. C. (2010). 'Empirical studies in industrial location: An assessment of their methods and results', Journal of Regional Science, 50, pp. 685-711.

Arauzo-Carod, J. M. and Viladecans-Marsal, E. (2009). 'Industrial location at the intra-metropolitan level: The role of agglomeration economies', Regional Studies, 43, pp. 545-558.

Autant-Bernard, C. (2006). 'Where do firms choose to locate their R&D? A spatial conditional logit analysis on French data', European Planning Studies, 14, pp. 1187-1208.

Bade, F. J. and Nerlinger, E. A. (2000). 'The spatial distribution of new technology-based firms: Empirical results for West-Germany', Papers in Regional Science, 79, 155-176.

Barrios, S., Görg, H. and Strobl, E. (2006). 'Multinationals' location choice, agglomeration economies, and public incentives', International Regional Science Review, 29, pp. 81–107.

Bellak, C., Leibrecht, M. and Riedl, A. (2008). 'Labour costs and FDI flows into Central and Eastern European countries: A survey of the literature and empirical evidence', Structural Change and Economic Dynamics, 19, pp. 17–37.

Broadman, H. G. and Sun, X. (1997). 'The distribution of foreign direct investment in China', World Economy, 20, pp. 339–361.

Cameron, A. C. and Trivedi, P. K. (2009). Microeconometrics Using Stata. Vol. 5. College Station: Stata Press.

Cameron, A. C., Trivedi, P. K. (1986). 'Econometric models based on count data: Comparisons and application of some estimators and tests', Journal of Applied Econometrics, 1, 29-54.

Carlin, W., Van Reenen, J. and Wolfe, T. (1995). 'Enterprise restructuring in early transition: The case study evidence from central and eastern Europe', Economics of Transition, 3, pp. 427–458.

Carlton, D. (1983). 'The location and employment choices of new firms: an econometric model with discrete and continuous endogenous variables', Review of Economics and Statistics, 65, pp. 440–449.

Cheng, S. and Stough, R. R. (2006). 'Location decisions of Japanese new manufacturing plants in China: a discrete-choice analysis', Annals of Regional Science, 40, pp. 369-387.

Chidlow, A., Salciuviene, L. and Young, S. (2009). 'Regional determinants of inward FDI distribution in Poland', International Business Review, 18, pp. 119–133.

Chung, W. and Alcácer, J. (2002). 'Knowledge seeking and location choice of foreign direct investment in the United States', Management Science, 48, pp. 1534–1554.

Cieślik, A. (2005a). 'Regional characteristics and the location of foreign firms within Poland', Applied Economics, 37, pp. 863–874.

Cieślik, A. (2005 b). 'Location of foreign firms and national border effects: The case of Poland', Tijdschrift voor Economische en Sociale Geografie, 96, pp. 287-297.

Cieślik, A. (2013). 'Determinants of the location of foreign firms in Polish regions: Does firm size matter?', Tijdschrift voor Economische en Sociale Geografie, 104, pp. 175-193.

Cieślik, A. and Rokicki, B. (2013). 'Rola sieci transportowej w rozwoju polskich regionow: zastosowanie modelu potencjału ekonomicznego', Acta Universitatis Nicolai Copernici, Ekonomia, 44, pp. 113-127.

Coughlin, C. C, Terza, J. V, and Arromdee, V. (1991). 'State characteristics and the location of foreign direct investment within the United States', Review of Economics and Statistics, 73, pp. 675–683.

Coughlin, C. C. and Segev, E. (2000). 'Location determinants of new foreign-owned manufacturing plants', Journal of Regional Science, 40, pp. 323–351.

Crozet, M., Mayer, T. and Mucchielli, J-L. (2004). 'How do firms agglomerate? A study of FDI in France', Regional Science and Urban Economics, 34, pp. 27-54.

Egeln, J., Gottschalk, S. and Rammer, C. (2004). 'Location decisions of spin-offs from public research institutions', Industry and Innovation, 11, pp. 207-223.

Fidrmuc, J. P. (2007). 'Channels of restructuring in privatized Czech companies', Economics of Transition, 15, pp. 309–339.

Figueiredo, O., Guimarães, P. and Woodward, D. P. (2002). 'Home-field advantage: location decisions of Portuguese entrepreneurs', Journal of Urban Economics, 52, pp. 341–361.

Gauselmann, A. and Marek, P. (2011). 'Regional determinants of MNE's location choice in transition economies', WIFO Working Paper No. 412.

Gehringer, A. (2013). 'Knowledge externalities and sectoral interdependences: Evidence from an open economy perspective', University of Göttingen, mimeo.

Glaeser, E. (1999). 'Learning in cities', Journal of Urban Economics, 46, pp. 254-277.

Glickman, N. J. and Woodward, D. P. (1988). 'The location of foreign direct investment in the United States: patterns and determinants', International Regional Science Review, 11, pp. 137–154.

Goldstein, G. and Gronberg, T. (1984). 'Economies of scope and economies of agglomeration', Journal of Urban Economics, 16, pp. 91-104.

Greene, W. (2007). 'Fixed and random effects models for count data', Working Paper, Department of Economics, Stern School of Business, New York University.

Guimarães, P., Figueiredo, O. and Woodward, D. (2000). 'Agglomeration and the location of foreign direct investment in Portugal', Journal of Urban Economics, 47, pp. 115–135.

GUS (Glowny Urzad Statystyczny) (2013). 'Wpływ potencjału demograficznego i gospodarczego miast wojewódzkich na kondycję województw', Warsaw Statistical Office.

Head, K. and Ries, J. (1996). 'Inter-city competition for foreign investment: static and dynamic effects of China's incentive areas', Journal of Urban Economics,40, pp. 38–60.

Head, K. C., Ries, J. C. and Swenson, D. L. (1995). 'Agglomeration benefits and location choice: evidence from Japanese manufacturing investments in the United States', Journal of International Economics, 38, pp. 223–247.

Helsley, R.W. and Strange, W.C. (1990). 'Agglomeration economies and matching in a system of cities', Regional Science and Urban Economics, 20, pp. 189-212.

Hilber, C. and Voicu, I. (2010). 'Agglomeration economies and the location of foreign direct investment: Empirical evidence from Romania', Regional Studies, 44, pp. 355–371.

Holl, A. (2004). 'Start-ups and relocations: Manufacturing plant location in Portugal', Papers in Regional Science, 83, pp. 649-668.

Krugman, P. (1991). 'Increasing returns and economic geography', Journal of Political Economy, 99, 483-499.

Krugman, P. (1996). 'Urban concentration: the role of increasing returns and transport costs', International Regional Science Review, 19, pp. 5–30.

Luger, M. and Shetty, S. (1985). 'Determinants of foreign plant start-ups in the United States: lessons for policy makers in the Southeast', Vanderbilt Journal of Transnational Law, 18, pp. 223–245.

Marshall, A. (1890). Principles of Economics, London: Macmillan.

Mudambi, R. and Swift, T. (2012). 'Multinational enterprises and the geographical clustering of innovation', Industry and Innovation, 19, pp. 1-21.

Quigley, J. (1998). 'Urban diversity and economic growth', Journal of Economic Perspectives, 12, pp. 127-138.

Robinson, I. (2004). 'Enterprise restructuring in transition: Case study evidence from Poland', Journal of East-West Business, 10, pp. 45-64.

Rosenthal, S. S. and Strange, W. C. (2001). 'The determinants of agglomeration', Journal of Urban Economics, 50, pp. 191-229.

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

Woodward, D. P. (2000). 'Locational determinants of Japanese manufacturing start-ups in the United States', Southern Economic Journal, 58, pp. 690–708.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data, MA: MIT Press.

Appendix

Table A.1: List of variables

Variable	Description
Firms, total	Entities of the economy entered, entities of the national economy-indicators, NTS-5, entities entered in the *Regon* register per 10 thousand population
Firms, new	Entities of the economy entered, entities of the national economy-indicators, NTS-5, new entities of the national economy recorded in the *Regon* register per 10 thousand population
Price index	Price indices of consumer goods and services, total, NTS-2
Population	Population, Population and vital statistics, Population by domicile/residence and sex, NTS-5, total locations, actual place of residence, as of 31 December, males and females
GDP	Gross domestic product total, current prices, PKD 2007, NTS-2
Urban population	Population by domicile/residence and sex, in urban areas, actual place of residence, as of 31st December, total, NTS 5
Industry agglomeration	Economic activity of the population (average annual data), Structure of employed persons by economic sector and sex, NTS-2, industry sector, total
Services agglomeration	Economic activity of the population (average annual data), Structure of employed persons by economic sector and sex, NTS-2, services sector, total
Wages	Wages and salaries and social security benefits, wages and salaries, average monthly gross wages and salaries in national economy by PKD 2004, NTS-2, and for lagged value 2010 by NACE rev. 2, NTS-2
Students	Higher education, higher education institutions, higher education institutions by type, NTS-3, total, students
Unemployment rate	Labour market, Economic activity of the population (average annual data), unemployment rate by place of residence, NTS-2, total
Telephone lines	Transport and communication, Communication, Main telephone lines of all operators, in total, NTS-2, total locations, total subscribers
Road network	Transport and communication, Public roads, roads-indicators, NTS-2, roads ward surface per 100 km²
Railway lines	Transport and communication, Rail transport, railway lines-indicators, NTS-2, railway lines standard gauge per 100 km²
Area	Territorial division, Geodetic area, Area, NTS-5, total in ha, 2009
R&D agglomeration economies	Science and technology, Research and development activity, Employment in R&D-indices, NTS-2, employed persons per 1000 economically active persons
Patents per labor workforce	Patents applications to the EPO by priority year and NTS 3 regions (pat_ep_rtot), per million labor force, from EUROSTAT
German-Polish border	Distance on one of three main routes for cars from the regional capital cities to the German border cities Penkun, Görlitz or Frankfurt Oder, in km, given by measured distances through google maps
EU	EU membership dummy which counts 1 in 2004 and the years thereafter

Table A.2: Summary statistics

Variable		Mean	Std. Dev.	Min	Max	Observations
Number of all	overall	230507.4	142212.4	87409.2	681032.4	N = 128
registered firms	between		145844.7	90752.6	622165.8	n = 16
	within		11228.5	184057.2	289374.0	T = 8
Number of newly	overall	18831.4	11615.3	5047.3	66365.1	N = 128
registered firms	between		11257.9	6418.3	47714.8	n = 16
	within		3893.9	9197.0	37481.7	T = 8
Population	overall	2385034.0	1204879.0	1008196.0	5222167.0	N = 128
	between		1239421.0	1008920.0	5169367.0	n = 16
	within		15485.8	2344290.0	2437834.0	T = 8
GDP	overall	63064.2	51283.4	18030.0	273612.5	N = 128
	between		50979.5	22634.0	215692.7	n = 16
	within		13203.6	15114.5	120984.0	T = 8
Urban population	overall	59.5	10.0	40.4	79.0	N = 128
in percent	between		10.3	40.6	78.6	n = 16
	within		0.3	58.5	60.3	T = 8
Industry	overall	29.7	5.3	18.1	40.7	N = 128
agglomeration	between		5.2	20.1	39.5	n = 16
economies	within		1.7	26.2	34.6	T = 8
Services	overall	52.8	6.0	41.0	65.1	N = 128
agglomeration	between		5.9	44.3	62.4	n = 16
economies	within		1.7	48.2	57.6	T = 8
Wages	overall	2239.8	358.6	1788.1	3644.6	N = 128
	between		261.6	2039.1	3091.7	n = 16
	within		252.8	1850.1	2792.7	T = 8
Students per	overall	4.7	0.9	2.8	6.8	N = 128
population	between		0.9	3.6	6.7	n = 16
	within		0.2	3.8	5.2	T = 8
Unemployment	overall	14.8	5.7	5.5	26.3	N = 128
rate	between		2.1	12.0	18.6	n = 16
	within		5.4	4.7	24.5	T = 8
Telephone lines	overall	27.9	4.9	18.3	39.1	N = 128
	between		3.2	22.9	34.9	n = 16
	within		3.8	19.9	34.4	T = 8
Road network	overall	85.7	30.5	50.5	170.6	N = 128
	between		31.2	51.2	164.4	n = 16
	within		2.4	77.7	94.2	T = 8
Railway lines	overall	6.9	3.2	3.3	18.9	N = 128
	between		3.2	3.6	17.7	n = 16
	within		0.2	6.4	8.1	T = 8
Area	overall	19542.5	6830.9	9509.3	36688.8	N = 128
	between		7027.4	9509.3	36688.8	n = 16
	within		0.0	19542.5	19542.5	T = 8
R&D	overall	0.4	0.2	0.1	1.1	N = 128
agglomeration	between		0.2	0.1	1.0	n = 16
economies	within		0.0	0.2	0.5	T = 8
Patents per labor	overall	8.3	5.9	0.2	28.5	N = 124
workforce	between		4.0	1.8	15.5	n = 16
	within		4.5	-0.3	22.8	T-bar = 7.75
German-Polish	overall	371.3	188.2	29.5	686.0	N = 128
border	between		193.6	29.5	686.0	N = 128
	within		0.0	371.3	371.3	n = 16
EU membership	overall	0.8	0.4	0.0	1.0	T = 8
	between		0.0	0.8	0.8	N = 128
	within		0.4	0.0	1.0	n = 16