TIGER

TRANSFORMATION, INTEGRATION and GLOBALIZATION ECONOMIC RESEARCH
CENTRUM BADAWCZE TRANSFORMACJI, INTEGRACJI I GLOBALIZACJI

TIGER Working Paper Series

No. 127

Income Disparities in Poland through the Lens of Economic Geography

Jesús López-Rodríguez

Warsaw, June 2012

Income Disparities in Poland through the Lens of Economic Geography

Jesús López-Rodríguez[1]

ABSTRACT

Second nature geography variables are very relevant in the explanation of income disparities across regions within countries and across countries. This paper uses the framework of the New Economic Geography to derive the structural equation which relates nominal wages with a distance weighted sum of incomes in the surrounding locations, the so called nominal wage equation. The estimation of this equation was carried out for 2000, 2004 and 2008 using regional data on Polish regions. The results of the analysis show that market access plays an important role in the explanation of income disparities in Poland. Moreover, the effects of market access have undergone several robustness tests to control for endogeneity issues linked to the construction of the market access variable and also to control for problems arising from shocks linked to spatially correlated but intertemporally uncorrelated omitted variables. Finally, using data for serveral years show us that the picture which emerges from the estimations is that the effect of market access on GDP per head levels is increasing over time leading to the conclusion that the core-periphery patterns in Poland have been reinforced ever since the first year of our analysis (2000).

Key Words: New Economic Geography, Market Access, Income Disparities, Poland

JEL Classification: R11, R12, R13, R14, F12, F23

[1] Universidad de A Coruña, Department of Economic Analysis and Business Administration, Faculty of Economics, Campus de Elviña, s/n. C.P.: 15071, A Coruña, Spain, Phone: +34 981 167050, ext: 2451, Fax: +34 981 167070; C+D Jean Monnet Group on Competition and Development, email: jelopez@udc.es.

1. Introduction

Economic activities tend to cluster at many geographical levels (Florence (1948), Hoover (1948), Fuchs (1962), Enright (1990), Ellison and Glaeser (1996), Dumais *et al*. (1997), Porter (1998, 2000). At a world wide level, there is the so called North-South dualism where NAFTA (North American Free Trade Agreement- US-Canada-Mexico) the EU-15 countries and East Asia accounted for 83% of the world's GDP in the year 2000. Furthermore, Hall and Jones (1999) observe that high-income nations are clustered in small cores in the Northern hemisphere and that productivity per capita steadily declines with increasing distance from the core regions (New York, Brussels and Tokyo). In the case of Europe, a set of adjacent regions stretching from southeast England, through the Benelux countries, North France and Southwest Germany to Northeast Italy has been denoted the "Blue Banana". This area represents roughly 20% of the former EU15, but contains 40% of its GDP and 50% of its population (see Lopez-Rodriguez *et al*. 2007). This geographic concentration of economic activities can be seen also at national level. Examples in the case of Europe are for instance France, where the region of Île-de-France (the metropolitan area of Paris) accounts for 2.2% of the area of the country and 18.9% of its population and produces 30% of its GDP; Spain where the Northeast area plus Madrid represents a quarter of the total Spanish Peninsular area but concentrates almost 50% of its population and 60% of its GDP[2]. Poland is not an exception to this general trend we observe in the countries mentioned above. Polish regions such as Mazowieckie, where the capital of the country is, Slaskie, Dolnoslaskie and Wielkopolskie have per capita GDP levels well above the country´s average and are much more developed than Podkarpackie, Lubelskie and Podlaskie located in the far East part of the country along the axis North-South.

Table 1 shows the evolution of the GDP per capita across the sixteen NUTS 2 regions in which Poland is divided[3] (data is in the Polish national currency (zloty)) for the years 2000, 2004 and 2008). The results show quite clearly the dominance of the country´s

[2] There are more cases in Europe with this core-periphery pattern such as Portugal, UK, etc.

[3] Comparisons of per capita monetary amounts can be distorted if we do not use a relative price index for each region. However as the wage equation of the model will be defined in nominal terms we have decided to present in table 1 the current values of the variables instead of weighting them by a relative price index for each region.

capital, Warsaw, in terms of GDP per capita over the three periods of time under the analysis. Mazowieckie is by far the richest region in the country, being its GDP per capital 63% higher than the average GDP per capita income in 2000, 65% higher in 2004 and 72% higher in 2008.

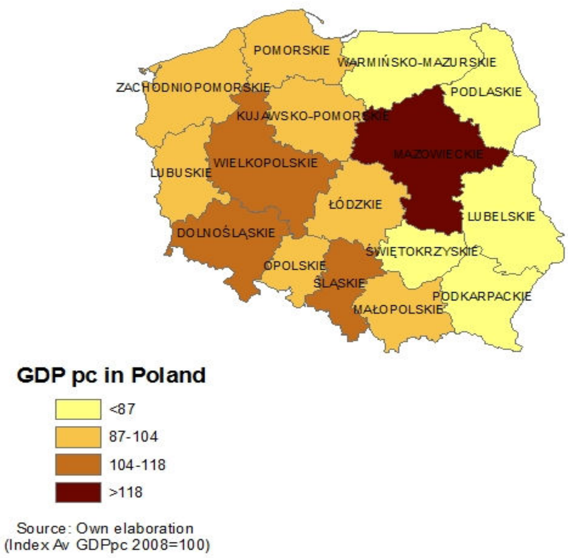
However the most important feature about the evolution of GDP per capita disparities in Poland lies on the fact that the income gap between Mazowieckie and the average GDP per capita of the country has been widening over time increasing from 2000 to 2008 nine percentage points. The data also shows that the gap between the richest and the poorest region has been widening over time. GDP per capita in Mazowieckie was 216% higher than in the poorest region in 2000, increasing the gap to 220% in 2004 and 229% in 2008.

Table 1: GDP per capita across NUTS 2 Regions (National Currency zloty)

Regions	GDP pc 2000	GDP pc 2004	GDP pc 2008
MAZOWIECKIE	29382	36884	52837
ŁÓDZKIE	17511	22261	31088
MAŁOPOLSKIE	16936	20664	28908
DOLNOŚLĄSKIE	20222	24641	35969
WIELKOPOLSKIE	20717	26011	34979
POMORSKIE	19332	23619	31813
ZACHODNIOPOMORSKIE	19793	22456	30359
KUJAWSKO-POMORSKIE	17803	21625	28930
LUBELSKIE	13620	16771	23189
ŚLĄSKIE	20963	27205	36086
PODLASKIE	14642	18053	24419
OPOLSKIE	16430	20779	28313
LUBUSKIE	17528	21623	28713
PODKARPACKIE	13574	16886	23101
WARMIŃSKO-MAZURSKIE	15239	18790	24818
ŚWIĘTOKRZYSKIE	15083	18696	26723
Av. GDP pc	18048	22310	30640
Highest GDP pc	29382	36884	52837
Lowest GDP pc	13574	16771	23101
Highest/Av. GDP pc	1,63	1,65	1,72
Lowest GDP pc /Av. GDP pc	0,75	0,75	0,75
Highest GDP pc/Lowest GD	2,16	2,20	2,29
Source: Authors Own calculations			

Moreover, these disparities show a well-defined "core-periphery" type of structure in the sense that we can see a clear spatial income gradient for the Polish GDP pc. Map 1 plots GDP pc across Polish NUTS 2 regions relative to the country´s 2008 average. From the map it is clear the Western-Eastern division of the country in terms of income being the Western part performing as the *core economic areas* and the Eastern as the *periphery economic areas* in a New Economic Geography sense. This pattern can also be seen in figure 1 which plots GDP per capita against distance to Wroslaw, more or less the geographical centre in terms of income. The results show that as we move further away from Wroslaw, GDP per capita (on average) decreases.

Map 1: GDP per capita across Polish Regions in 2008



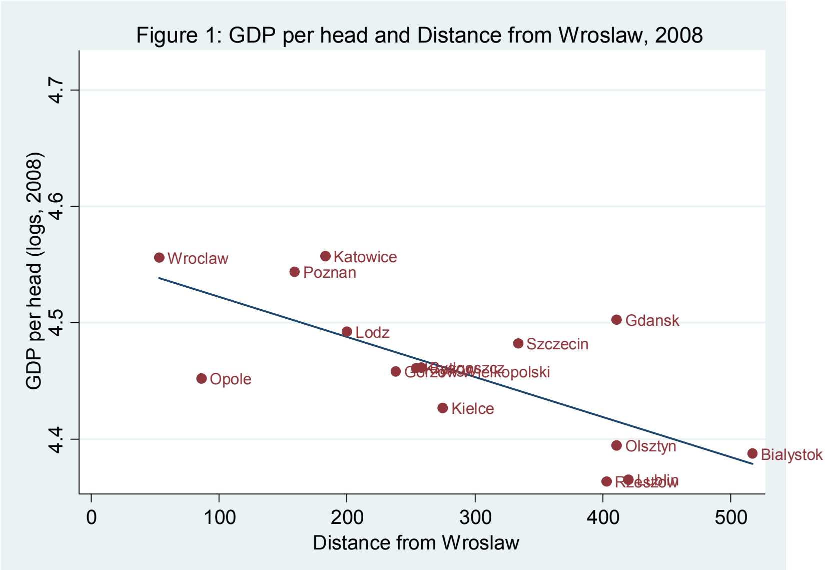


Figure 1: GDP per head and Distance from Wroslaw, 2008

Source: Own elaboration based on data from Eurostat and Cambridge Econometrics databases

Differences in economic performance across regions or countries can be explained by a myriad of different theories such as economic growth theories, economic development theories, urban economics theories, etc.. These theories will prompt us to different underlying factors in economic performance such as differences in savings rates, investment rates, skilled human capital and difficulties in technology transmission (Barro and Sala-i-Martin, 1995), natural advantages of different locations, such as access to navigable rivers, ports, airports, allocation of oil, sunshine (Gallap et. al 2000, Hall and Jones, 1999). However we will resort to the so called *Geographical Economics* or *New Economic Geography* (Krugman (1991a, 1991b) as our main theoretical framework to explain income disparities across Polish regions. The main reason to resort to this alternative explanation is due to the fact that we think is a more satisfactory way to explain the agglomeration of economic activities than the explanations based on arguments of the economic growth literature or arguments based on *first nature* geography. The explanation of the differences in economic development within the New Economic Geography literature are based on the so called *second nature geography factors* which in the literature are usually refer to as market access or market potential[4]. Therefore, the New Economic Geography (NEG) literature contrary

[4] Market access in a particular location is a weighted sum of the volume of economic activities in the surrounding locations where the weighted scheme is the transportation costs between the different locations

to the economic growth literature explicitly takes into account the economic-geographical features of income and production.

This paper extends our knowledge on how market potential or market access affects the economic development across Polish regions over the period 2000 to 2008. First and most important, we show that the cross-sectional striking success of the economic geography to predict income per capita in Breinlich (2006), Hanson (2005), Lopez-Rodriguez and Faiña (2007) and Redding and Venables (2004), holds when considering cross-regional data for the case of Poland. This reinforces their findings strongly, and confirms the theoretical predictions of standard core-periphery New Economic Geography models for different settings. Second, the results are robust to an instrumentation strategy intended to capture endogeneity issues linked to the construction of the market access variable and to control for problems arising from shocks linked to spatially correlated but intertemporally uncorrelated omitted variables.

The remaining part of the paper is structured as follows: Section 2 gives the microeconomic foundations of the market access concept and establishes the relationship between second nature geography (approximated by market access) and the income level of a particular location, which in the literature of the New Economic Geography is known as the nominal wage equation. Section 3 contains the econometric specifications, data source and construction of the variables. Section 4 presents the results of the baseline estimations. Section 5 disentangles the channels of influence by looking at potential third variables that might be affecting per capita GDP levels and are related to market access and finally, section 6 contains a summary of the main contributions of the paper.

2. New Economic Geography, market access and the nominal wage equation

Our theoretical framework is a reduced form of a standard New Economic Geography model (multiregional version of Krugman 1991b model) [5] which incorporates the key ingredients to obtain the so called nominal wage equation which will constitute the workhorse of our empirical estimation.

[5] Other related NEG models can be seen in Fujita et al. (1999). See also López-Rodríguez and Faiña (2008) for a detailed explanation of the building blocks of a NEG model and for the full derivation of a 2x2x2 canonical model.

We consider a world with R regions (j =1, 2……….R), and we focus on the manufacturing sector, composed of firms that produce a great number of varieties of a differentiated good (D) under increasing returns to scale and monopolistic competition. Transportation costs of differentiated goods are in the form of Iceberg costs so in order to receive 1 unit of the differentiated good in location j from location i, $T_{i,j} > 1$ units must be shipped, so $T_{i,j} = 1$ means that the trade is costless, while $T_{i,j} - 1$ measures the proportion of output lost in shipping from i to j . The manufacturing sector can produce in different locations.

On the demand side, the final demand in location j can be obtained by the Utility maximization of the following CES function:

$$\max_{m_{i,j}(z)} D_j \qquad (1)$$

Where D_j represents the consumption of the differentiated good in location j . D is an aggregate of the different industrial varieties defined by a CES function à la Dixit and Stiglitz (1977):

$$D_j = \left[\sum_{i=1}^{R} \int_0^{n_i} m_{i,j}(z)^{\sigma-1/\sigma} dz \right]^{\sigma/\sigma-1} \qquad (2)$$

where $m_{i,j}(z)$ represents the consumption of each variety z in location j and which is produced in location i , n_i is the number of varieties produced on location i , σ is the elasticity of substitution between any two varieties where $\sigma > 1$. If varieties are homogenous σ goes to infinite and if varieties are very different σ takes a value close to 1. Consumers maximize their utility (function #1) subject the following budget constraint:

$$\sum_{i=1}^{R} n_i x_{ij}^D p_{ij} = Y_j \qquad (3)$$

Solving the consumer optimization problem, we obtain the final demand in location j of each variety produced in location i.

$$x_{ij}^D = p_{ij}^{-\sigma} \left[\sum_{n=1}^{R} n_n p_{nj}^{1-\sigma} \right]^{-1} Y_j \qquad (4)$$

where p_{ij} ($p_{ij} = p_i T_{ij}$), is the price of varieties produce in location i and sold in j and Y_j represents the total income of location j.

If we define a price index for manufacturing goods as $P_j = \left[\sum_{n=1}^{R} n_n p_{nj}^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$

This Industrial Price index of location j measures the minimum cost of buying 1 unit of the differentiated good D so it can be interpreted as an expenditure function. If we rewrite the expenditure on consumption as $E_j = Y_j$ the final demand in location j can be given by $x_{ij}^{consD} = p_{ij}^{-\sigma} P_j^{\sigma-1} E_j$. However, in order for x_{ij}^{consD} units to arrive to location j, $T_{i,j} x_{ij}^{consD}$ units must be shipped. Thus effective demand facing a firm in i from j is given by expression:

$$x_{ij}^{D} = T_{ij} p_{ij}^{-\sigma} P_j^{\sigma-1} E_j = p_i^{-\sigma} T_{ij}^{1-\sigma} P_j^{\sigma-1} E_j \qquad (5)$$

Turning to the supply side, a representative country i firm maximizes the following profit function:

$$\Pi_i = \sum_{j=1}^{R} \frac{p_{ij} x_{ij}^{D}}{T_{i,j}} - w_i^{D} (F + c x_i^{D}) \qquad (6)$$

The technology of the increasing returns to scale sector is given by the usual linear cost function: $l_{Dij} = F + c x_{ij}^{D}$, where l_{Dij}, represents the industrial labor force needed to manufacture 1 unit in location i and sell it in location j, F, are the fixed costs units which are needed for manufacturing the industrial good, c, is the unit variable cost and x_{ij}^{D} is the quantity of each variety demanded in location j and produced in location i ($x_i^{D} \equiv \sum_j x_{ij}^{D}$ represents the total output produced by the firm in location i and sold in the different j locations) and w_i^{D} is the nominal wage paid to the manufacturing sector workers in location i. Increasing returns to scale, consumers ′love of variety and the existence of an limited number of access varieties of the manufacturing good mean that each variety is going to be produced by a single firm in a single location. In this way the number of manufacturing firms coincides with the number of varieties. Each firm maximizes is own profit behaving as a monopolist of its own variety of the differentiated good. First order conditions for profit maximization lead us to the standard result that prices are a mark-up over marginal costs.

$$p_i = \frac{\sigma}{\sigma - 1} w_i^D c \tag{7}$$

where $\frac{\sigma}{\sigma-1}$ represents the Marshall-Lerner Price-cost ratio. The higher this ratio, the higher the monopolistic power of the firm. Krugman (1991b) interprets σ as an inverse measure of the scale economies due to it can be interpreted as a direct measure of the price distortion and as an indirect measure of the market distortion due to the monopoly power. Due to $\frac{\sigma}{\sigma-1}$ is higher than 1, Krugman (1991b) interprets this result as a proof of increasing returns to scale. Substituting this pricing rule into the profit function we obtain the following expression for the equilibrium profit function:

$$\Pi_i = \left(w_i^D\right)\left[\frac{cx_i^D}{\sigma - 1} - F\right] \tag{8}$$

Free entry assures that long-run profits will be zero implying that no firm will have incentives to move from one location to another. This implies that equilibrium output is the following one:

$$x_i^D = \bar{x} = \frac{F(\sigma - 1)}{c} \tag{9}$$

The price needed to sell this many units is given by $P_i^{\sigma} = \frac{1}{\bar{x}} \sum_{j=1}^{R} E_j P_j^{\sigma-1} T_{i,j}^{1-\sigma}$.

Combining this expression with the fact that in equilibrium prices are a constant mark-up over marginal costs we obtain the following zero-profit condition:

$$w_i^D = \left(\frac{\sigma - 1}{\sigma c}\right)\left[\frac{1}{\bar{x}} \sum_{j=1}^{R} E_j P_j^{\sigma-1} T_{i,j}^{1-\sigma}\right]^{1/\sigma} \tag{10}$$

This equation is called nominal wage equation which constitutes the key relationship to be tested in the empirical part of this work. According to equation (10) the nominal wage level in each location i depends on a weighted sum of the purchasing capacities of the different j locations where the weighted scheme is a decreasing function of the distance between locations. In the New Economic Geography literature the expression on the right hand side of equation (10) has been labelled with different names market access (Redding and Venables (2001, 2004)) and real market access (see Head and Mayer (2004))[6].

We will refer to this expression as market access and will be labelled as (MA). The meaning of this equation is that access advantages raise local factor prices. More precisely, production sites with good access to major markets because of its relatively low trade costs tend to reward their production factors with higher wages.

If we normalize the way we measure production, choosing the units such as that $c = \frac{(\sigma - 1)}{\sigma}$, $F = \frac{1}{\sigma}$, and defining the market access of location i as $MA_i = \sum_{j=1}^{R} E_j P_j^{\sigma-1} T_{i,j}^{1-\sigma}$, we can rewrite the nominal wage equation as:

$$w_i^D = \left[MA_i\right]^{1/\sigma} \qquad (11)$$

This simplification of the nominal wage equation is very similar to the Harris (1954) market access function in the sense that economic activity is more important in those regions which are close to big markets.

3. Econometric Specification

Taking logarithms in expression (11), the estimated nominal wage equation is based on the estimation of the following expression:

$$\log(w_i) = \theta + \sigma^{-1} \log\left[MA_i\right] + \eta_i \qquad (12)$$

Where η_i is the error term and the other variables are as defined in the previous sections. This equation relates the nominal wage in region i with income in other regions, weighted by distance. Therefore, in accordance with the predictions of the

[6] This expression is semantically analogous to the one employed by Harris (1954) but the term real refers to the fact that price difference between different locations are taken into account. The concept of nominal market access of Head and Mayer (2004) is a concept similar to the Harris (1954) market access.

theory, the higher the levels of income and price levels and the lower the distance between locations, the greater the level of local wages. This specification captures the notion of a spatial wage structure and allows us to verify the direct relationship between the nominal wage of a location and market access which is an important condition for us to observe agglomeration dynamics.

However equation (12) is a restricted specification to analyze the access effects on wages of market access as we cannot say whether the regression captures the causality or simply captures correlations with omitted variables such as infrastructure, human capital, innovation levels and so on. To address these potential impacts and control for the possibility of other shocks that are affecting the dependent variable and are correlated with market access, we also estimate an alternative specification that explicitly takes into account the above considerations. The extended estimation of the nominal wage equation takes the following form:

$$Logw_{\mathrm{i}} = \theta + \sigma^{-1} \ln MA_i + \sum_{n=1}^{N} \gamma_n X_{i,n} + \eta_i \qquad (13)$$

Where X_{in} is a vector of control variables and γ_{in} the corresponding coefficient.

3.1 Data Source and Construction of variables

The data we use in our work refers to the years 2000, 2004 and 2008. The data was taken from different sources: The data on NUTS 3 regions gross domestic product (GDP) and population was taken from the Polish Office for National Statistics. GDP data is measured at current national prices (zloty). Data on R&D expenditure and human capital comes from Eurostat and Cambridge Econometrics databases. With respect to the market access variable we build a distance-weighted sum of the volume of economic activity in the surrounding regions whereby the weighting scheme is a function declining with increasing distance between locations i and j. The distance between places i and j is measured in kilometres between the capital cities of the regions. Therefore we build a market access variable for each year of our analysis.

The calculation of distances in the market access variable is made on the basis of the 66 regions in which Poland is divided at NUTS 3 level. To calculate the distance between regions we take the road distance between their capital cities expressed in kilometres.

For the calculation of the internal distance within each region, it is approximated by a function that is proportional to the square root of each prefecture area. The expression used for calculation is $0.66\sqrt{\frac{Area}{\pi}}$ where "Area" represents the size of the region expressed in km2. This expression gives the average distance between two points on a circular location (see Crozet 2004, Head and Mayer, 2000, and Nitsch 2000) for a discussion of this measure of internal distance).

As control variables in our work we have decided to incorporate R&D expenditures at regional level since they might be affecting the GDP per capita values through market access.

4. Empirical Results

4.1 Market Access and Income: Econometric Estimations

In this section we carry out cross-section estimations of equation 12 based on data on GDP per head for the years 2000, 2004 and 2008 and two different measures of market access computed also for the years 2000, 2004 and 2008. Table 2 summarizes our baseline results. In columns 1 and 2 we regress GDP per head across Polish NUTS3 regions on two different measures of market access. In Column 1 we use as the input to compute market access the gross domestic product of each region and in column 2 the population in each region is used instead. Columns 3 and 4 precede in the same way for the year 2004 and columns 5 and 6 for the year 2008. These first set of results show that the different coefficients of market access across the six estimations are positive and statistically significant at the usual significant levels and therefore the results are in line with the theoretical predictions of the model. Using the first measure of market access, doubling the market access of a region would increase GDP per head between 47% and 57% while with our second measure doubling the market access of a region would increase GDP per head between 35% and 47%. The explanatory power of the regressions varies between 27% and 33% with our first measure of market access and between 16% and 20% with our second. From a dynamic perspective, these results suggest that the effects of market access on income levels are getting stronger and stronger over time, or in other words the core-periphery structure of Poland is reinforcing over time. This result is in line with a very detailed analysis on the evolution in market potential values carried out by Lopez-Rodriguez and Runiewizc-Wardyn (2011).

Table 2: Market Access and Regional Income: Baseline Estimations

Dependent Variable	Log GDPpc 2000		Log GDPpc 2004		Log GDPpc 2008	
Regressors	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.55** (0.70)	2.34** (0.88)	2.46** (0.72)	2.14** (0.93)	2.30** (0.79)	1.97** (0.90)
MAGDP2000	0.47** (0.19)					
MAPOP2000		0.35** (0.16)				
MAGDP2004			0.51** (0.19)			
MAPOP2004				0.41** (0.17)		
MAGDP2008					0.57** (0.21)	
MAPOP2008						0.47** (0.17)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS
R2	0.27	0.16	0.29	0.17	0.33	0.20
Prob (F-statistic)	0.00	0.00	0.00	0.00	0.00	0.00
Number observations	66	66	66	66	66	66

Note: Table displays coefficients for OLS estimations and Huber-White heterocedasticity robust standard errors in parenthesis. The dependent variable is the log of a region's gross domestic product per head in the years 2000, 2004 and 2008. LogMAGDP2000, LogMAGDP2004 and LogMAGDP2008 are the logs of market access in the years 2000, 2004 and 2008 computed based on GDP figures. LogPOP2000, LogMAPOP2004 and LogMAPOP2008 are the logs of market access in the years 2000, 2004 and 2008 computed based on population figures (see text for details of calculation of LogMA). For data sources see text and appendix A. * and ** signify statistical significance at the 5% and 1% levels

4.2 Robustness Checks

In this section we are going to carry out several robustness checks in the baseline estimation (equation 12). In first place potential shocks in GDP per head will be captured by our disturbance term and therefore they will be correlated across regions. This fact brings us the problem of correlation with our market access measure since market access in each region is a distance weighted sum of GDPs in the surrounding ones. In order to address this concern columns 1 to 6 of table 3 regress GDP per head across NUTS 3 Polish regions on lagged values of market access (Columns 1 and 2 use one year lag in market access) and columns 3, 4 , 5 and 6 use four year lags in market access. This estimation avoids problems arising from shocks linked to spatially correlated but intertemporally uncorrelated omitted variables. The results obtained in these estimations are very much in line with those obtained in table 2.

Table 3: Robustness Checks (lagged values)

Dependent Variable	Log GDPpc 2001		Log GDPpc 2004		Log GDPpc 2008	
Regressors	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.63** (0.74)	2.49** (0.91)	2.47** (0.70)	2.17** (0.92)	2.41** (0.73)	2.01** (0.89)
LogMAGDP2000	0.45** (0.21)					
LogMAPOP2000		0.33** (0.17)				
LogMAGDP2000			0.52** (0.20)			
LogMAPOP2000				0.40** (0.17)		
LogMAGDP2004					0.56** (0.20)	
LogMAPOP2004						0.46** (0.17)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS
R2	0.24	0.12	0.29	0.17	0.33	0.20
Prob (F-statistic)	0.00	0.00	0.00	0.00	0.00	0.00
Number observations	66	66	66	66	66	66

Note: Table displays OLS coefficients and Huber-White heterocedasticity robust standard errors in parenthesis. The dependent variable is the log of a region's gross domestic product per head in 2001, 2004 and 2008. LogMAGDP2000 and LogMAGDP2004 are the logs of market access in the years 2000 and 2004 computed based on GDP figures. LogPOP2000 and LogMAPOP2004 are the logs of market access in the years 2000 and 2004 computed based on population figures (see text for details of calculation; column 1 and 2 use LogMA lagged by one period). Columns 3, 4, 5 and 6 use LogMA lagged four periods. For data sources see appendix A. * and ** signify statistical significance at the 5% and 1% levels.

One concern that arises in the estimation of the effect of market Access on GDP per head is that market access is an endogenous regressor. Market access includes GDP in its computations which in turn are increasing in GDP per head. As a way to deal with this problem we carry out our estimations using instrumental variables. These variables need to generate a variation in gross domestic product only through their impact on market access. Taking into account these premises and following other studies carried out on spatial economic issues linked to the nature of this research (Lopez-Rodriguez et al. (2007b) and Redding and Venables (2004)), in this paper we use as instruments geographical variables. These are the most suitable candidates for such estimation and are exogenous determinants of market access. We first instrument market access with distance from Wroslaw in Km (Dist Wroslaw in the table). The first instrument captures the market access advantages of locations close to the geographic centre of Poland. However one can argue that Wroslaw is a centroid of GDP per head´s distribution within Poland and distance to it might capture other determinants of income levels besides market access. Therefore we also instrument market access with the average distance each region is to the rest of regions in the country measured in km (Avg Dist in the table). This second instrument shows a more direct link to the channel modelled in the theoretical section of this paper—which stresses trade cost savings of more central or better linked regions. These instruments need to pass two tests: the "first stage" restriction, which tests whether the variation in the instrument is correlated to the variation in the endogenous variable –in this case, market access-, and the exclusion restriction, which cannot be tested empirically.

Formally, we can represent the Two –Stage Least Square estimation we are going to implement in the following way:

$$\ln \hat{MA_{\mathrm{i}}} = \theta + \beta Z_i + \sum_{n=1}^{N} \gamma_n X_{i,n} + \varepsilon_i \qquad (14)$$

$$LnGDPperhead_{\mathrm{i}} = \theta + \sigma^{-1} \ln M\hat{A}_i + \sum_{n=1}^{N} \gamma_n X_{i,n} + \eta_i \qquad (15)$$

Where MA is the endogenous regressor, on the outcome LnGDPperhead and Z is the instrument we are going to use. In the same way, we can represent the aforementioned restrictions:

- First Stage Restriction: $\beta \neq 0$
- Exclusion Restriction: $\mathrm{cov}(Z_i, \eta_i) = 0$

Table 4 reports the results for the cross sectional regression model for the years 2000, 2004 and 2008 using both instrumental variables described above. The instruments are highly statistically significant and have the expected signs. The p-value for an F-test of the null hypothesis that the coefficients on the excluded instruments are equal to zero is 0.00. Distance to Wroslaw explains 61% of regional market access and average distance of a region to other regions 73%. Columns 2, 4 and 6 regress GDP per head on market access for the years 2000, 2004 and 2008 using our first instrument (Distance to Wroslaw) and in columns 8, 10, and 12 we rerun the same regressions instrumenting market access with average distance. The results of the estimations show that the coefficients on market access are positive and statistically significant according to the usual significance levels though a bit lower in the regressions that instrument market access with average distance. However comparing the results in table 5 with our baseline estimations, the market access coefficients increase substantially when IV estimation is carried out. Doubling a region´s market access would increase GDP per head by between (56%-66%-Instrument Avg Dist) or between (60%-71%-Instrument Dist Wroslaw) compare to a range between (47%-57%) in the baseline estimations. Once again another interesting result that emerges from the IV estimations is that the effect of market access on GDP per head levels is increasing over time leading to the conclusion that the core-periphery structure in Poland has been reinforced ever since the year 2000.

Table 4: Robustness Checks (IVE)

Dependent Variable Regressors	Log MAGDP2000 (1)	Log GDPpc2000 (2)	Log MAGDP2004 (3)	Log GDPpc2004 (4)	Log MAGDP2008 (5)	Log GDPpc2008 (6)	Log MAGDP2000 (7)	Log GDPpc2000 (8)	Log MAGDP2004 (9)	Log GDPpc2004 (10)	Log MAGDP2008 (11)	Log GDPpc2008 (12)
Constant		2.05** (0.89)		1.95** (0.88)		1.74** (1.02)		2.23** (0.26)		2.13** (0.97)		1.95 (1.00)
Dist Wroslaw	-0.0005** (0. .0001)		-0.0005** (0.0001		-0.0005** (0.0001)							
Avg Dist							-0.001** (0. .0001)		0..001** (0.0001)		-0.0018** (0.0001)	
logMAGDP2000		0.60** (0.25)						0.56** (0.27)				
LogMAGDP2004				0.64** (0.17)						0.60** (0.26)		
LogMAGDP2008						0.71** (0.26)						0.66** (0.26)
Estimation	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
R2	0.61	0.13	0.61	0.13	0.61	0.14	0.73	0.27	0.73	0.28	0.75	0.32
F-Stat and (p-value) based on excluded instruments	18.77 (0.00)	0.00	18.02 (0.00)	0.00	19.04 (0.00)	0.00	91.54 (0.00)	0.00	97.89 (0.00)	0.00	105.57 (0.00)	0.00
Hansen J-Stat (p-value)	Exactly indentified		Exactly indentified		Exactly indentified		Exactly indentified		Exactly indentified		Exactly indentified	
Number observations	66	66	66	66	66	66	66	66	66	66	66	66

Note: Table displays coefficients and t-statistics for IV estimation based on Huber–White robust standard errors. The dependent variable in columns 2, 4, 6, 8, 10 and 12 is the log of a region's gross domestic product per head for the years 2000, 2004 and 2008 respectively. The independent variable in these columns is the log of market access for the years 2000, 2004 and 2008 respectively (LogMAGDP2000, LogMAGDP2004, LogMAGDP2008 see text for details of calculation). Instruments for logMAGDP2000, logMAGDP2004 and logMAGDP2008 used are distance to Wroslaw (in Km) (Columns 1, 3 and 5) and average distance to other regions (Avg Dist) (Columns 7, 9 and 11).Columns 1, 3, 5, 7 and 11display the corresponding first-stage results. For data sources see appendix A. * and ** signify statistical significance at the 5% and 1% levels. Table displays coefficients and Huber-White heterocedasticity robust standard errors in parenthesis.

5. Disentangling Channels of Influence

The analysis carried out in the previous sections shows basically two things: On the one hand there is a well-defined spatial income structure across Polish regions where higher market access regions enjoy also higher levels of GDP per head and on the other hand the results also point to a reinforcement of this spatial structure over time due to the fact that market access coefficients are higher at the end of the our period of analysis than at the beginning. However the results of these bivariate regressions between GDP per head and market access may be due to third variables that are affecting regional Polish GDP per head levels through the market access and work through accumulation incentives such as for instance the case of the expenditure in research and development. There are several papers that deal with the relationship between innovative activities and their link to spatial proximity and geography (see Bilbao-Osorio and Rodriguez-Pose (2004), Bottazzi and Peri (1999, 2003), Moreno et al. (2005) and Rodriguez-Pose (1999, 2001)).

Table 5 shows the regression results of R&D expenditure on market access for the years 2000, 2004 and 2008. Although a through testing for the determinants of R&D in Poland is not the goal of this paper, the results in the table point out also to a spatial structure in the distribution of R&D expenditure across Polish regions which is quite linked to regional market access values. Based on these results a way to unravel the effects of market access on GDP per head levels will be incorporating R&D expenditure into the analysis as an additional regressor. If including R&D expenditure into the regression keeps the coefficient on market access positive and statistically significant would mean that the market access variable is indeed important in the determination of income levels in Poland.

Table 5: Market Access and Research and Development: 2000, 2004, 2008

Dependent Variable Regressors	Log R&D2000 (1)	Log R&D2004 (2)	Log R&D2008 (3)
MAGDP2000	2.92** (1.09)		
MAGDP2004		2.89** (1.22)	
MAGDP2008			3.18** (1.21)
Estimation	OLS	OLS	OLS
R2	0.28	0.24	0.30
Prob (F-statistic)	0.00	0.00	0.00
Number observations	66	66	66

Note: Table displays OLS coefficients and Huber-White heterocedasticity robust standard errors in parenthesis. The dependent variables are the log of a region´s research and development expenditure for the years 2000, 2004 and 2008 (columns 1, 2 and 3) and the share (in % points) of the working populations with tertiary educational attainment levels for the years 2000, 2004 and 2008 (columns 4, 5 and 6 respectively). For data sources see text and appendix A.
** indicates coefficient significant at 0.05 level, *indicates coefficient significant at 0.01 level. Results on the included constant are suppressed

Table 6: Disentangling Channels of Influence: Static Analysis

Dependent Variable Regressors	Log GDPpc2004 (1)	Log GDPpc2004 (2)	Log GDPpc2008 (3)	Log GDPpc2008 (4)
Constant	2.52** (0.75)	2.92** (0.56)	2.48** (0.77)	3.03** (0.51)
LogMAGDP2004	0.43** (0.20)			0.31** (0.13)
LogR&D2004	0.11 (0.06)	0.12** (0. .05)		0.33** (0.14)
LogMAGDP2000		0.33** (0.14)		
LogMAGDP2008			0.45** (0.20)	
LogR&D2008			0.12 (0.07)	0.14 (0. .07)
Estimation	IV	OLS	IV	OLS
R2	0.52	0.54	0.54	0.56
F-Stat and (p-value) based on excluded instruments	18.77 (0.00)	0.00	18.02 (0.00)	0.00
Hansen J-Stat	Exactly indentified		Exactly indentified	
Number observations	66	66	66	66

Note: Table displays coefficients and t-statistics for IV and OLS estimations based on Huber–White robust standard errors (patenthesis). The dependent variable in columns 1 and 2, is the log of a region's gross domestic product per head for the year 2004 and in columns 3 and 4 is the log of a region's gross domestic product per head for the year 2008. The independent variables in these columns are the log of market access for the years 2004 (columns 1 and 4) 2000 for column 2 and 2008 for column 3 (see text for further details of calculation). Instruments for logMAGDP2004 and logMAGDP2008 used are average distance to other regions (Avg Dist) (Columns1 and 3). Columns 1 and 2 use as additional regressor Log R%D2004 and 3 and 4 Log R%D2008
For data sources see appendix A. * and ** signify statistical significance at the 5% and 1% levels.

We have performed an static analysis (table 6) and a dynamic analysis (table 7) of the combined effects of market access and R&D expenditure on GDP per head levels across Polish Regions. In the static analysis (table 6) we regress the GDP per head of 2004 against market access 2004 and R&D expenditure 2004) using IV estimation (Column 1) and against market access lagged 4 years and R&D expenditure 2004. In columns 3 and 4 we repeat the process for the year 2008 using GDP per head in 2008 as our dependent variable and market access in 2008 and R&D expenditure 2008 as our independent variables using IV estimation (Column 3) and lagging market access four periods (Column 4). The results show that although there is a reduction on the market access coefficient values they are still positive and statistically significant at the usual critical levels. R&D expenditure is statistically significant only in the second and fourth regression.

Table 7: Disentangling Channels of Influence (I): Dynamic Analysis (2000-2004)

Dependent Variable Regressors	Log GDPpc2004 (1)	Log GDPpc2004 (2)	Log GDPpc2004 (3)	Log GDPpc2004 (4)	Log GDPpc2004 (5)	Log GDPpc2004 (6)	Log GDPpc2004 (7)	Log GDPpc2004 (8)	Log GDPpc2004 (9)	Log GDPpc2004 (10)
Constant	2.52**	2.50**	2.54**	2.51**	2.75**	2.92**	2.90**	2.89**	2.87**	3.06**
	(0.75)	(0.79)	(0.80)	(0.76)	(0.78)	(0.56)	(0.57)	(0.58)	(0.58)	(0.58)
LogMAGDP2004	0.43**	0.43**	0.42	0.43**	0.35					
	(0.20)	(0.21)	(0.21)	(0.20)	(0.22)					
LogMAGDP2000						0.33**	0.33**	0.33**	0.34**	0.26
						(0.14)	(0.15)	(0.15)	(0.15)	(0.16)
LogR&D 2004	0.11					0.12**				
	(0.06)					(0.05)				
LogR&D 2003		0.12**					0.13**			
		(0.05)					(0.05)			
LogR&D 2002			0.12**					0.13**		
			(0.05))					(0.05)		
LogR&D 2001				0.12					0.13**	
				(0.06)					(0.06)	
LogR&D 2000					0.15**					0.16**
					(0.05))					(0.05)
Estimation	IV	IV	IV	IV	IV	OLS	OLS	OLS	OLS	OLS
R2	0.52	0.56	0.54	0.53	0.61	0.54	0.57	0.55	0.54	0.61
F-Stat and (p-value) based on excluded instruments	97.89 (0.00)	97.89 (0.00)	97.89 (0.00)	97.89 (0.00)	97.89 (0.00)					
Hansen J-Stat	Exactly indentified	Exactly indentified	Exactly indentified	Exactly indentified	Exactly indentified					
Number observations	66	66	66	66	66	66	66	66	66	66

Note: Table displays coefficients and t-statistics for IV and OLS estimation based on Huber–White robust standard errors. The dependent variable is the log of a region's gross domestic product per head for the year, 2004.. The independent variable in these columns is the log of market access for the year 2004 (columns 1 to 5) and 2000 (columns 6 to 10) and the Log R&D for 2004 (Columns 1 and 6) and then yearly lags of Log R&D to 2000 (Columns 2 to 5 and 7 to 10) (see text for details. Instruments for logMAGDP2004 and logMAGDP2000 used are average distance to other regions (Avg Dist) (Columns 1 to 6)
Columns 1 to 6 3, 5, 7 display the corresponding first-stage results. For data sources see appendix A. * and ** signify statistical significance at the 5% and 1% levels.

Table 8: Disentangling Channels of Influence (II): Dynamic Analysis (2004-2008)

Dependent Variable Regressors	Log GDPpc2008 (1)	Log GDPpc2008 (2)	Log GDPpc2008 (3)	Log GDPpc2008 (4)	Log GDPpc2008 (5)	Log GDPpc2008 (6)	Log GDPpc2008 (7)	Log GDPpc2008 (8)	Log GDPpc2008 (9)	Log GDPpc2008 (10)
Constant	2.48** (0.77)	2.38** (0.80)	2.31** (0.81)	2.26** (0.74)	2.38** (0.76)	3.03** (0.51)	2.90** (0.54)	2.87** (0.54)	2.79** (0.54)	.53** (0.53)
LogMAGDP2008	0.45 ** (0.20)	0.48 ** (0.21)	0.51 (0.21)	0.51** (0.19)	0.49** (0.19)					
LogMAGDP2004						0.31** (0.13)	0.35** (0.14)	0.37** (0.13)	0.39** (0.13)	0.36** (0.13)
LogR&D 2008	0.12 (0.07)					0.14 (0.07)				
LogR&D 2007		0.11 (0.06)					0.13** (0.06)			
LogR&D 2006			0.11 (0.06)					0.12** (0.06)		
LogR&D 2005				0.12** (0.05)					0.13** (0.05)	
LogR&D 2004					0.11 (0.06)					0.13** (0.05)
Estimation	IV	IV	IV	IV	IV	OLS	OLS	OLS	OLS	OLS
R2	0.54	0.53	0.54	0.59	0.57	0.56	0.55	0.55	0.60	0.58
F-Stat and (p-value) based on excluded instruments	105.57 (0.00)	105.57 (0.00)	105.57 (0.00)	105.57 (0.00)	105.57 (0.00)					
Hansen J-Stat	Exactly indentified	Exactly indentified	Exactly indentified	Exactly indentified	Exactly indentified					
Number observations	66	66	66	66	66	66	66	66	66	66

Note: Table displays coefficients and t-statistics for IV and OLS estimation based on Huber–White robust standard errors. The dependent variable is the log of a region's gross domestic product per head for the year, 2008.. The independent variable in these columns is the log of market access for the year 2008 (columns 1 to 5) and 2004 (columns 6 to 10) and the Log R&D for 2008 (Columns 1 and 6) and then yearly lags of Log R&D to 2004 (Columns 2 to 5 and 7 to 10) (see text for details. Instruments for logMAGDP2004 and logMAGDP2000 used are average distance to other regions (Avg Dist) (Columns 1 to 6)
Columns 1 to 6 3, 5, 7 display the corresponding first-stage results. For data sources see appendix A. * and ** signify statistical significance at the 5% and 1% levels.

Table 7 and 8 summarize the results of the dynamic analysis in which we extended the estimations of the equation (equation 13) using as time frames 2000-2004 (table 7) and 2004-2008 (table 8). In order to control for the endogeneity problems arising in the estimations we run and report the results of the IV estimations (Columns 1 to 5) in both tables instrumenting market access with our preferred instrument, average distance of a region to the rest of regions. In order to avoid problems arising from shocks linked to spatially correlated but intertemporally uncorrelated omitted variables we have also performed OLS estimations using four year lags in our market access variable (columns 6 to 10 in both tables). Market access in 2000 is used instead its value in 2004 in columns 6 to 10 of table 7 and market access in 2004 is used instead its 2008 value in columns 6 to 10 of table 8. With respect to the R&D expenditure variable the so called dynamic approach in tables 7 and 8 comes with the idea that we resort to annual lags in R&D expenditure values which will give a dynamic image of the association between these lags and GDP per head levels. The intuitive idea of using R&D lags is based on the fact that R&D expenditure in one year will have a forward impact in GDP per head levels if these expenditures are transformed into product or process innovations, new patents, etc.

Our main focus is again on the significance of the market access coefficient in the different set of regressions we carried out. The results, except a few cases, show that market access coefficients are positive and statistically significant and tend to remain so despite the introduction of different lags in the R&D variable. The exceptions are in the year 2004 in the IV estimations when two year and four year lag in the R&D expenditure are used (Column 3 and 6) and also in the OLS estimation and four year lag in the R&D variable. Regarding the estimations for the year 2008, only the IV estimation with two lags in R&D (Column 3) does not show up as significant. Once again the impact of market access in GDP per head levels is higher under IV estimations than under OLS and on average between 50% and 60% of the spatial variation in GDP per head levels can be explained by including information on market access and R&D expenditure levels.

Finally, the results of these set of regressions confirms the results of the theoretical model showing that market access is an important variable when we analyze the difference in GDP per head levels across Polish NUTS 3 regions. Moreover our results point out to the fact that expenditure in R&D might be playing an important role in the shape of income levels across Polish regions.

6. Conclusions

This paper estimates one of the structural equations derived from core-periphery new economic geography models, the so called nominal wage equation, using data from the Polish regions for three periods of time, 2000, 2004 and 2008. The main results of the estimations are in line with the theoretical predictions of the model, showing that there is a spatial structure of income levels across Polish Regions. The results of the baseline estimation of GDP per head against market access have shown that market access plays an important role in the explanation of GDP per head disparities in Poland. The baseline results have undergone several robustness tests to control for endogeneity problems by means of IV estimations and also to control for problems arising from shocks linked to spatially correlated but intertemporally uncorrelated omitted variables performing OLS estimations using year lags for our market access variable. The results of these alternative estimations have also shown that market access is still positive and statistically significant in the explanation of GDP per head levels across Polish regions.

Additionally, in order to unravel the effects of variables that work through accumulation incentives and which therefore could be linked to market access we have extended the baseline estimations incorporating as a control in the estimations the shares of expenditure in R&D at regional level. The results of the estimations have also shown that market access continues to be significant.

Moreover, the use of data for three different years during the 2000 thousands in our estimations allows us to see how the market access coefficients have behaved during these years. An interesting result that emerges from the estimations is that the effect of market access on GDP per head levels is increasing over time leading to the conclusion that the core-periphery patterns in Poland have been reinforced ever since the first year of our analysis (2000). This fact calls for policy actions to counterbalance the tendency towards the agglomeration of economy activities in the core regions in Poland. At this respect the EU regional policy should play a big role in even out the development levels across Polish regions on account of not only the natural tendency of concentration of economic activities in space but in this case the continuous reinforcement of the core-economic Polish regions.

Acknowledgements

A significant part of this paper has been finished while the author was a Visiting Scholar in the department of economics at Harvard University during 2012. The author also wants to thank very especially professors Pol Antràs and Nathan Nunn for sponsoring his visit to Harvard University. Financial support from the Spanish Ministry of Science and Innovation through Grant ECO2011-28632 is gratefully acknowledged. The usual disclaimer applies

7. References

Barro R.J. and Sala-i-Martin X. (1995) Economic Growth, New York, McGraw-Hill.

Bottazzi, L., and Peri, G. (1999): "Innovation, Demand and Knowledge Spillovers: Theory and Evidence From European Regions", CEPR Discussion Papers 2279.

Bottazzi, L. and Peri, G., (2003): "Innovation and spillovers in regions: Evidence from European patent data" European Economic Review, 47, 687-710.

Bilbao-Osorio, B. and Rodriguez-Pose A. (2004): "From R&D to Innovation and Economic Growth in the EU", Growth and Change, 35, 434-455.

Breinlich, H. (2006): "The Spatial Income Structure in the European Union - What Role for Economic Geography?", Journal of Economic Geography, 6, 593–617

Crozet M. (2004): "Do Migrants Follow Market Accesss? An Estimation of a New Economic Geography Model", Journal of Economic Geography , 4, 439-458.

Dumais, G, Ellison, G and Glaeser, E. L (2002) Geographic concentration as a dynamic process, Review of Economics and Statistics, 84:193-204.

Ellison, G and Glaeser, E (1997) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach, Journal of Political Economy 105: 889-927.

Enright, M (1990) Geographic concentration and industrial organization. PhD thesis, Harvard University.

Fuchs, V (1962) Changes in the location of manufacturing in the US since 1929. New Haven: Yale University press

Fujita M., P. Krugman and A. Venables (1999), The Spatial Economy (MIT Press, Cambridge MA).

Florence, P. S. (1948) Investment, Location and size of plant, London: Cambridge University press.

Gallup J., Mellinger A. and Sachs J. (2000) Climate, Coastal Proximity and Development, Chapter 9. In: Clark, G., Gertler, M., Feldman, M. (Eds), The Oxford Handbook of Economic Geography. Oxford Univ. Press, Oxford, pp. 169-194

Hall, R.E. and C.I Jones (1999) Why do some countries produce so much more output per worker than others? Quarterly Journal of Economics 114, 83-116.

Hanson, G. (2005): "Market access, increasing returns and geographic concentration", Journal of International Economics, 67, 1-24.

Harris, C. (1954), The market as a factor in the localization of industry in the United States, Annals of the Association of American Geographers 64, pp. 315–348.

Head K. and Mayer T. (2004), The Empirics of Agglomeration and Trade, in Henderson V. and J.F.Thisse (eds.) Handbook of Urban and Regional Economics, Vol. 4, Chapter 59, pp 2609-2669

Head, K.; Mayer, T. (2000): Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU, Weltwirtschaftliches Archiv 136, 284-314.

Krugman, P. (1991a), Geography and Trade (MIT Press, Cambridge MA).

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

López-Rodríguez J. and Faiña A. (2007) Regional Wage Disparities in the European Union: What role for Market Access, Investigaciones Regionales, 11, pp.5-23.

López-Rodríguez J., Faiña A. (2008): "New Economic Geography: Empirical Evidence of the Estimation of the Nominal Wage Equation," Lecturas de Economía, 69, 41-62

Lopez-Rodriguez and Runiewizc-Wardyn M. (2011) Changing Market Potentials, Regional Growth and Income Disparities in Poland, Working paper (unpublished)

López-Rodríguez, J, Faína, A and García Lorenzo, A (2007a) The Geographic Concentration of Population and Income in Europe: Results for the Period 1984-1999, Economics Bulletin, 18:1-7.

López-Rodríguez J., Faiña J.A. and Lopez Rodriguez J. (2007b): "Human Capital Accumulation and Geography: Empirical Evidence from the European Union", Regional Studies, 42, 217-234.

Moreno, R., Paci, R. and Usai, S. (2005): "Spatial spillovers and innovation activity in European regions", Environment and Planning A, 37(10), 1793 – 1812.

Nitsch V. (2000) National borders and international trade: evidence from the European Union, Can J Econ 22:1091–1105

Porter, M (2000) Location, Competition, and Economic Development: Local Clusters in a Global Economy, Economic Development Quarterly, 14:15-34.

Porter, M (1998) Clusters and the New Economics of Competition, Harvard Business Review, November-December.

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

Rodriguez-Pose, A. (1999): “Innovation prone and innovation averse societies. Economic Performance in Europe”, Growth and Change 30, 75-105.

Rodriguez-Pose, A. (2001): “Is R&D investment in lagging areas of Europe worthwhile? Theory and Empirical evidence”, Papers in Regional Science 80, 275-295.