Agglomeration Economies: a Bridge between the New Economic Geography and the Malthusian Theory

M. Agovino
University of Naples "Parthenope"
Faculty of Economics
Department of Economic Studies
Via Medina, 40 - 80133 Naples (Italy)
agovino.massimo@gmail.com

A.Garofalo
University of Naples "Parthenope"
Faculty of Economics
Department of Economic Studies
Via Medina, 40 - 80133 Naples (Italy)
antonio.garofalo@uniparthenope.it
Phone: (+) 39 081 5474778
Fax: (+) 39 081 5474750

Abstract

The aim of this paper is to build a bridge between the New Economic Geography and the Malthusian theory. Both theories seek to explain the phenomenon of agglomeration processes, yet along two different lines: Malthusian theory through differences in geographical factors, historical accidents, cultural and social factors and so on (*first nature effects*); the New Economic Geography through scale economies or knowledge spillovers (*second nature effects*). Based on this assumption, we use a methodology based on an **AN**alysis **O**f **Va**riance (**ANOVA**) proposed by Roos (2005) which we apply to a panel data of Italian Local Labour Systems (LLS) for the years 2001-2005. We are thus able to quantify how much of GDP can be derived from natural geographical circumstances (first nature elements) and how much from agglomeration economies or second nature elements (in our case, population and GDP). Our results reveal that although the *gross second nature* effect plays a major role in bringing about the agglomeration process, the *net second nature* effect appears relatively unimportant. By contrast, both *first nature effects* and the *combined effect* of the two forces are very important.

JEL: *R12, O52, N10, N30, N50, O10, 040.*

Keywords: *Agglomeration, New Economic Geography, Malthusian Theory*

1. Introduction

Population increases over time are generally explained by the fact that improvements in economic conditions have led, roughly speaking, to a reduction in mortality rates and finally to a reduction in birth rates. As is widely recognised, a good proxy of economic conditions is income per capita since it reflects, among other things, the impact of technology, education and health. The usual explanations therefore suggest that there is a strong link between income per capita and population. At the same time, however, where these variables have a high value we can speak of agglomeration. The relationship between the two variables (namely income per capita and population) changes according to the theory considered. More precisely, according to the first regime known as the *Malthusian regime* the relationship between per capita income and population growth is positive, where small increases in income lead to an increase in population growth. In the second stage, also known as *post-Malthusian regime*, the relationship between income and population growth remains positive. In the final stage, the *modern growth regime*, there is a rapid growth in per capita income whereas population growth declines. As a result, there is a negative relationship between the two variables (per capita income and population).

Our goal in this paper is not to verify the existence (or otherwise) of the relationship between population and per capita income but rather to verify the impact of these variables in determining agglomeration processes. In other words, we seek to investigate the causes of the so-called agglomeration economies; for this purpose we will use both the population and GDP since they allow us to control for the second nature effects (agglomeration economies).

Both the New Economic Geography (in one way) and the Malthusian Theory (in an indirect way) consider agglomeration processes which may arise in some geographic areas rather than in others. Malthusian theory is able to explain such processes with reasons primarily related to geography, environment, history, and so on (first nature effects). By contrast, the New Economic Geography seeks to justify agglomeration processes introducing the concept of scale economies and spillover effects (second nature effects), which are not connected at all with the environmental and/or geographic contexts in which agglomeration processes occur. What is certain is that both factors are important in causing agglomeration processes.

Investigating the causes of different growth paths for different economic realities is difficult and entails controlling for numerous variables, which are rather difficult to observe. To this extent, Galor (2005) says: "*Variations in the economic performance across countries and regions…reflect initial differences in geographical factors and historical accidents and their manifestation in* *variations in institutional, demographic, and cultural factors, trade patterns, colonial status, and public policy.*"

It is important to recall that many factors affect the distribution of economic activity. Traditionally, characteristics linked to the physical landscape, such as temperature, rainfall, access to the sea, the presence of natural resources or the availability of arable land (also known as the first nature effects), are distinguished from factors relating to human actions and economic incentives, such as scale economies or knowledge spillovers (which belong to second nature effects).

According to first nature causes the agglomeration of firms and households can be explained by an accidental accumulation of favourable natural features. However, judging from second nature causes, agglomerations are due to *agglomeration economies*; the theory states that it must be advantageous for agents to be where many other agents are, irrespective of the particular geographic location. Although Krugman's theory (Krugman, 1993, 1999) shows that agglomerations can be explained by second nature causes alone, agglomeration is in fact caused by both first and second nature effects.

In this paper, we examine the influence of geographic features and agglomeration economies on the location of production in the Italian **L**ocal **L**abour **S**ystem. In other words, we focus on quantifying how much of the geographic pattern of GDP can be attributed to only exogenous first nature elements, how much can be derived from endogenous second nature factors and how much is due to the interaction of both effects. For this purpose we apply a methodological approach used by Roos (2005), who proposes using **AN**alysis **O**f **VA**riance (**ANOVA**) to infer the unobservable importance of first nature effects indirectly in a stepwise procedure. Aiming to disentangle first and second nature effects empirically, we control for second nature because every locational endowment will be reinforced and overlaid by second nature advantages.

The paper is organized as follows: Section 2 describes data and the methodology used. In Section 3, we present the results and related comments; Section 4 concludes.

2. Methodology and Data Description

Basically, there are three forces leading to agglomeration processes: an unobservable direct effect of the first nature, a first nature effect working through induced agglomeration economies and a direct effect of second nature, which would exist even without any first nature forces. In this paper, our goal is to quantify these forces; this will be done by using Roos's approach (2005), namely a method based on **ANOVA**. The total variance V of the dependent variable (in our case GDP per area) can be decomposed into four parts:

$$V = V_u + V_f + V_s + V_{fs} \qquad (\mathbf{1})$$

where V is the total variance of the dependent variable, V_u is the unexplained variance, V_f is the variance explained by first nature alone, V_s is the variance explained by second nature alone and V_{fs} is the variance explained by a combination of both forces. In order to better capture the methodology used, we report in ***Table 1*** a scheme which proposes the various steps of the approach. According to Malthus (1992), population grows whenever per capita income is above the subsistence level (Samuelson, 1988). This implies that there is a direct relation between per capita income and population size. An increase in income per capita leads to an increase in population (Faria et al. 2006); hence an increase in both variables results in a concentration of them in some places rather than in others. Based on this assumption, our goal will be to explain agglomeration from first and second nature elements. For this reason, we choose as a variable able to explain the agglomeration processes the relative GDP density -GDP per Km^2- (Roos, 2005; Chasco et al., 2008) while population and GDP per worker will be used as variables required to capture the second nature effects. Formally, the endogenous variable is defined as follows:

$$\log(gd_i) = \log \frac{Y_i / A_i}{\sum_i (Y_i / A_i)}$$

where Y_i is GDP and A_i is the area of region $\boldsymbol{i}$. The relative GDP density of a region is its GDP density relative to the average density of all regions. If $\log(gd_i)$ is equal to zero, region i's GDP share is equal to its area share; if it is larger (smaller) than zero, the region has a concentration of economic activity above (below) the average.

In this section, we explore the geographic dimension of GDP per area and population per area for the 686 Italian **L**ocal **L**abour **S**ystems[1] (**LLS**) during the period 2001-2005[2]. More precisely, we concentrate on the GDP per area for the whole country. We then disaggregate into five macro-areas, namely North, North-East, North-West, Centre and South.

From ***Figures 1*** and ***2*** we note that the highest average (and median) GDP per area is recorded for the South of Italy while the lowest is found in the North-West, for both years. The lowest variability

[1] Specifically, LLS are sets of contiguous municipalities with a high degree of self-containment of daily commuter travel. These areas can be considered the empirical counterpart of theoretical local production systems (LPS). LLSs are geographical units which, as opposed to regions and provinces, are not arbitrarily defined from an economic point of view. In fact, these units try to match the definition of labour markets as closely as possible in the following sense: they are made of spatially connected *comuni*; more than 75% of the residents in the unit work within the unit itself; firms mostly employ a local workforce.

[2] The source for these variables is ISTAT (National Institute of Statistics).

is recorded for the South of Italy while the largest occurs in the North-East. As regards population by area, the highest average population per area is found in the North-East of Italy, the lowest in the North-West. The highest population variability per area is observed in the North and North-East, the lowest in the South. These initial results are of interest since they are in line with the Modern Growth Regime which suggests an inverse relationship between population and GDP, which is particularly true for the South of Italy.

Table 1. Analysis of Variance (ANOVA): a summary of the methodology

Step	Objective	Methodology	Results
First	We filter gross second nature indicators from first nature interrelations	We regress two **gross second nature**[3] variables (population and GDP per worker) on first nature (natural endowment, physical geography, location and political geography) to explain how much of the gross second nature effects is caused by purely first nature. $\log(pop_i) = \gamma_0 + \sum_{i=1}^{K} \gamma_k f_{ki} + resPOP_i$ $\log(prod_i) = \rho_0 + \sum_{k=1}^{K} \rho_k f_{ki} + resPROD_i$ (2)	We extract the residuals that are our net second nature variables (**resPOP** and **resPROD**).
Second (a)	We estimate how much of GDP per area variance can be explained by **gross second nature advantages.**	We regress GDP per area , $\log(gd_i)$, which is a proxy variable to measure economic agglomeration, on two gross second nature variables: $\log(gd_i) = \alpha_0 + \phi_1 \log(pop_i) + \phi_2 \log(prod_i) + \varepsilon_i$ (3)	Determinant coefficient (R^2) indicates this gross effect of second nature on agglomeration: $R^2_{gs} = \frac{(V_s + V_{fs})}{V}$
Second (b)	We estimate how much of GDP per area variance can be explained by **net second nature advantages**	We regress GDP per area , $\log(gd_i)$, on two net second nature variables (the residuals that we previously extracted): $\log(gd_i) = \alpha_0 + \phi_3 resPOP_i + \phi_4 resPROD_i + \varepsilon_i$ (4)	Determinant coefficient (R^2) indicates this net effect of second nature on agglomeration: $R^2_{ns} = \frac{V_s}{V}$
Intermediate step	We extract the mixed effect of the interaction between first and second nature on GDP density.	We make the difference between the determinant coefficient of the second(a) step and that of the second(b) step.	$\frac{V_{fs}}{V} = R^2_{gs} - R^2_{ns}$

[3] Since first and second nature are interrelated, in this step we filter out the second nature variables (population and GDP per worker) empirically by first nature effects. In this way, we obtain a net second nature effects (which are represented by the residues extracted from the regression).

Third	We estimate how much of GDP per area variance can be explained **jointly by gross first and second nature**	We regress GDP per area , $\log(gd_i)$, on two net second nature variables and on first nature variables: $\log(gd_i) = \alpha_0 + \phi_3 resPOP_i + \phi_4 resPROD_i + \sum_{i=1}^{K} \lambda_k f_{ki} + \varepsilon_i$ (5)	Determinant coefficient (R^2) indicates the joint importance of first and second nature: $R^2_{f+s} = \frac{(V_s + V_{fs} + V_f)}{V}$
Fourth	We derive the difference between the results in the third and second step which is the importance of **first nature** alone.	We make the difference between the determinant coefficient of the third step and that of the second (a) step.	$\frac{V_f}{V} = R^2_{f+s} - R^2_{gs}$

Source: our elaboration on Roos' paper (2005)

Figure 1. Descriptive statistics of GDP per area and population per area (2001)

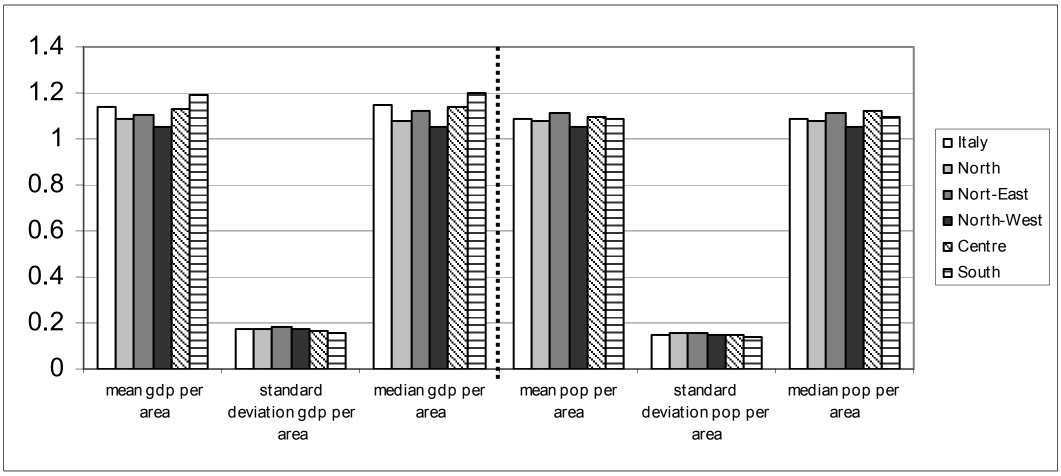
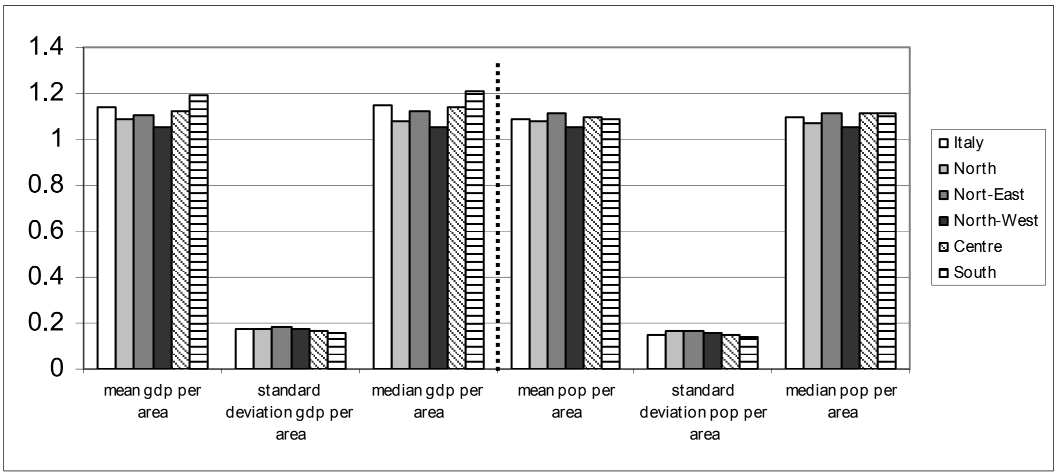


Figure 2. Descriptive statistics of GDP per area and population per area (2005)



We conclude this section by presenting some indicators able to measure first and second nature effects. As regards first nature effects, we are interested in those geographical characteristics related to the distribution of economic activity. This is the case of **natural endowment**, such as physical geography, relative location and political geography[4]. In order to capture those features, following Chasco et al. (2008), we chose the following indicators[5]: the annual average daily precipitation (*rain*) and a dummy variable for the presence of agricultural systems (*agric*), metro system as an indicator of air quality (*metro*). In this context, dummy variables are preferred to production quantities since our aim is to measure the exogenous endowment, not the endogenous output (Roos, 2005).

Furthermore, in line with Gallup and Sachs (1998), Rappaport (2000) and Limão and Venables (2001) we consider indicators of **physical geography** such as: altitude, seismicity, the presence of port systems- if they are classified as coastal communities- (coast) and three climate variables[6], namely the average annual maximum temperature (*tmax*), the average annual minimum temperature (*tmin*) and average annual solar radiation (*sunsh*). We also consider some variables related to **location**, namely latitude-longitude earth coordinates (*xcoo, ycoo*). Location is considered a significant geographical feature affecting agglomeration (Chasco et al., 2008).

Finally, we consider **political geography** variables which have been highlighted by several authors (Mathias, 1980; McCallum, 1995; Roos, 2005) who consider that agglomeration is positively or negatively affected by presence in the cluster of a capital city or by being a border region, respectively. Like Roos (2005) and Chasco et al. (2008), we consider regional capitals (*capreg*) since such places are those where legislative and executive power are usually concentrated and they have better access to information about regional government investments and planning decisions (Ades and Glaeser, 1995; Funck, 1995; Ayuda et al., 2005). We also consider a dummy for LLSs in border regions (*border*) due to the major differences in terms of language, culture and institutions existing between Italian LLSs and the European regions which they border.

As for the **second nature** variables, we follow Roos (1995) who uses total population (*pop*) and labour productivity (*prod*) since on aggregate levels both variables can capture many agglomeration economies (such as informational spillovers and labour market economies).

3. Results and comments

[4] Examples of **natural endowments** positively related to GDP density are agriculture, minerals, natural resources, and good soil and water supply (Gallup et al. 1999, Rapaport and Sachs 2003; Ayuda et al. 2005; Roos 2005).

[5] We report the description in the table in the **Appendix.**

[6] The source of climatic variables is the UCEA (Central Bureau of Agricultural Ecology). They were interpolated with kriging methodology since they are only available from a small number of monitoring stations.

In this section, we apply the methodology proposed by Roos (2005) for Germany and used by Chasco et al. (2008) for European countries, as summarized in ***Table 1***. We propose, in contrast to previous studies which suggested cross-section analysis, panel data analysis that considers 686 area units for the years 2001-2005.

Proceeding with the **first step,** we regress the second nature variables (population and GDP per worker) on geography and take the residuals as variables of net second nature forces (equation 2); in so doing we verify that the regressions of population and productivity on the 11 first nature variables lead to high multicollinearity. In order to overcome this problem, by following Chasco et al. (2008) we implement factor analysis[7] for the entire set of variables of the first nature. Interpretation of the three extracted factors is as follows: **Factor1** is a variable that takes account of location (latitude, longitude) and climate characteristics (rainfall, temperature, radiation); **Factor2** is mainly related to the physical characteristics (altitude), while **Factor3** is based on air quality. Table 2 presents the results of the final regressions of the second nature variables on first nature.

Table 2. Second nature on first nature

Dependent variable	Log(pop)						Log(prod)					
Variables	**Italy**	**North**	**North East**	**North West**	**Centre**	**South**	**Italy**	**North**	**North East**	**North West**	**Centre**	**South**
Factor1	-.0029*** (-3.98)	-.0027** (-2.35)	-.0114*** (-5.16)	.0019 (1.30)	-.0073** (-2.33)	-.0043*** (-4.59)	-.0086 (-1.25)	-.0389*** (-4.48)	-.1034*** (-6.12)	-.0033 (-0.31)	-.0026 (-0.13)	.0045 (0.39)
Factor2	.0152*** (5.05)	.0304*** (5.66)	.0493*** (5.73)	.0254*** (3.61)	-.0129* (-1.66)	.0189*** (4.68)	.0070 (0.25)	.2953*** (7.46)	.4410*** (6.76)	.2506*** (5.04)	-.2066*** (-4.07)	-.1329*** (-2.66)
Factor3	.0128*** (3.01)	.0446*** (5.15)	.0066 (0.652)	.0349*** (2.77)	-.0316*** (-2.90)	.00354 (0.68)	-.0890** (-2.25)	.2531*** (3.97)	-.0208 (-0.19)	.1765** (1.98)	-.2070*** (-2.91)	-.2775*** (-4.30)
Constant	4.5572*** (42231.66)	4.6645*** (2385.66)	4.5899*** (971.63)	4.7365*** (2284.60)	4.5367*** (3207.47)	4.4840*** (4519.17)	-3.1996*** (-3197.82)	-3.1330 *** (-218.12)	-3.3074*** (-92.32)	-3.0181*** (-206.47)	-3.1697*** (-343.44)	-3.3592*** (-274.24)
R^2	0.0072	0.0038	0.4814	0.0311	0.0816	0.1214	0.0160	0.0170	0.1112	0.0083	0.0042	0.0089
F	13.38***	20.01***	16.37***	13.70***	7.91***	10.62***	7.91***	24.30***	20.34***	13.54***	10.78***	8.99***
Hausman	567.76***	298.24***	199.34***	113.45***	90.42***	203.51***	310.47***	81.46***	51.31***	32.16***	20.63***	30.87***
observations	3430	1165	610	555	640	1625	3430	1165	610	555	640	1625

****, **,*: significant at: 1%, 5%, 10%; log(pop) is logarithm of population, log(prod) is logarithm of labour productivity.*
( ): t-test

The estimates presented in ***Table 2*** were obtained with the fixed effects method; this choice was made since *Hausman's test* always rejects the null hypothesis, preferring an estimate with fixed effects rather than one with random effects. The *F test* always rejects the null hypothesis of no joint significance of individual effects. Measured by R^2, the fit of both population and labour productivity equations is very low. We find a significant relation between second nature and geography only for the North-East of Italy while for the rest of the macro-areas, including the whole country, we experienced a very low level of interaction between the two forces. From estimation

[7] Factors were extracted using principal components and rotated with the Varimax method.

performed in ***Table 2*** we filtered the residuals, ***resPOP*** and ***resPROD***, which become our net second nature forces.

In proceeding, we encountered a problem reported by Roos (2005): the second nature variables are endogenous and simultaneously determined with GDP. This might lead to simultaneity bias in the regressions, does violate the necessary conditions to obtain estimates with good properties. Estimation of the instrumental variables (*IV*) is the standard approach to overcome the consequences of simultaneity.

We therefore proceeded to test the goodness of instruments to be used in equations (3), (4) and (5) with the *Sargan test*[8] and verify when the *IV* estimation should be implemented instead of ordinary least squares (*OLS*) estimation with the *Davidson-MacKinnon test*[9]. Instruments were chosen in line with Roos (2005), who proposed to use time-lagged variables, since they are highly correlated with the actual variables but also not at the same time correlated with the errors. ***Table 3*** lists the instruments used in each equation reported (the *Sargan test* and *Davidson-MacKinnon test).*

Table 3. Instruments and endogeneity tests

Equation	*Instrumented variables*	*Instruments*	*Sargan test*						*Davidson - MacKinnon test*					
			Italy	*North*	*North-East*	*North-West*	*Centre*	*South*	*Italy*	*North*	*North-East*	*North-West*	*Centre*	*South*
Gross second (equation 3)	Log(pop),Log(pod)	Log(pop)04,log(pop)03,log(pop)02, log(prod)04, log(prod)03, log(prod)02, resPOP05, resPOP04, resPOP03, resPOP02,resPROD05, resPROD04,resPROD03, resPROD02	14.509 (0.26)	10.368 (0.583)	9.780 (0.635)	9.140 (0.690)	16.383 (0.1743)	15.824 (0.1995)	3.3829**	4.3930**	1.7456	4.9340**	2.5966*	.3817
Net second (equation 4)	Pi, del	Log(pop)04,log(pop)03, log(pop)02, log(prod)04, log(prod)03, log(prod)02	7.440 (0.28)	6.390 (0.1719)	5.100 (0.2772)	3.435 (0.4879)	2.480 (0.6481)	4.880 (0.2998)	10.4611***	12.912***	6.8408**	.01338	1.1247	*8.2343****
First and Second Nature Joint Effect (Equation 5)	Pi, del	Log(pop)04,log(pop)03, log(pop)02, log(prod)04, log(prod)03, log(prod)02	1.400 (0.84)	3.696 (0.4487)	4.561 (0.3353)	6.086 (0.1928)	1.383 (0.8472)	7.665 (0.1047)	10.4298***	1.8028	1.9395	.3310	16.3766**	33.6793***

****, **,*: significant at: 1%, 5%, 10%; log(pop) is the logarithm of population, log(prod) is the logarithm of labour productivity, log(pop)xx is the logarithm of population for year xx, log(prod)xx is the logarithm for year xx, resPOPxx is the residual of the regression of the logarithm of population on first nature variables for year xx, resPRODxx resPOPxx is the residual of the regression of the logarithm of labour production on first nature variables for year xx*

In ( ): p-value

The *Sargan test* clearly accepts the null hypothesis confirming the validity of the instruments. The *Davidson-MacKinnon test* shows a high degree of simultaneity in almost all regressors with two exceptions in gross second nature (North-East and South), two in net second nature (North-West and Centre) and three in the first and second nature joint effect (North, North-East and North-West)[10].

[8] This test rejects the null hypothesis when at least one of the instruments is correlated with the error term (Sargan, 1964).

[9] The null hypothesis for which states that an ordinary least squares (OLS) estimator of the same equation would yield consistent estimates.[Syntax] A rejection of the null hypothesis indicates that the effects of endogenous regressors on the estimates are significant (Baum and Stillman, 2003*)*.

[10] These exceptions will be estimated with OLS.

Tables 4, 5 and ***6*** report the estimation results of equations *(3), (4)* and *(5)*. ***Table 4*** shows that the regressions on gross second nature variables provide a fairly significant determination coefficient (R^2_{gs})[11]although the largest value (about 37%) is reached in the North-East of Italy.

Regressions on net second nature variables **(*Table 5*)** provide a very low determination coefficient (R^2_{ns})[12], with the exception of the South of Italy which showed a value of just over 1%. This highlights the irrelevance of second nature effects in causing agglomeration processes for the *LLS*. Regarding the mixed effect of the interaction between first and second nature on GDP density (R^2_{fs}) (***Table 6***), a rather high percentage is observed for the North-East of Italy (62%), highlighting that the two forces work well together in leading to agglomeration processes in the LLS .

Table 4. Regression on gross second nature variables

Dependent variable: GDP per area						
Estimation	IV	IV	OLS	IV	IV	OLS
Variables	***Italy***	***North***	***North East***	***North West***	***Centre***	***South***
Log(pop)	.0747 (1.24)	.04919 (0.46)	.2835*** (6.25)	.2542** (1.97)	-.1295 (-0.75)	-.2546*** (-6.76)
Log(prod)	-.0681 *** (-8.35)	-.05202** (-2.37)	-.0995*** (-16.84)	-.0950*** (-4.11)	-.0500*** (-2.74)	-.0667*** (-21.86)
Constant	.5846** (2.08)	.69673 (1.29)	-.5058 ** (-2.31)	-.4331 (-0.67)	1.5593 * (1.91)	2.1052 *** (12.45)
R^2	0.2861	0.2459	0.3752	0.3527	0.2271	0.2901
Wald	1.02e+08***	3.10e+07***		1.60e+07***	1.85e+07***	
F			145.90***			265.22***
Observations	1372	466	610	222	256	1625

****, **,*: significant at: 1%, 5%, 10%; log(pop) is the logarithm of population, log(prod) is the logarithm of labour productivity*
In (): t-test

Table 5. Regression on net second nature variables

Dependent variable: GDP per area						
Estimation	IV	IV	IV	OLS	OLS	IV
Variables	***Italy***	***North***	***North East***	***North West***	***Centre***	***South***
Res POP	.0180897** (2.21)	.0111662* (1.70)	.0004105 (0.07)	.0007049 (0.09)	.0002592 (0.40)	-.0025895 (-0.53)
Res PROD	-.0424424** (-2.58)	.0060017 (1.00)	.0103938* (1.73)	.0002038 (0.02)	-.0002071 (-0.24)	-.0183513* (-1.89)
Constant	1.136845*** (638.97)	1.086345*** (995.16)	1.106906***	1.05631*** (492.09)	1.124935*** (3617.87)	1.188365*** (2137.50)
R^2	0.0003	0.0013	0.0029	0.0055	0.0028	0.0116
Wald	2.24e+07***	8.69e+06***	8.39e+06***			1.42e+07***
F				0.03	0.09	
Observations	1372	466	244	555	640	650

****, **,*: significant at: 1%, 5%, 10%; resPOP is residual of the regression of the logarithm of population on first nature variables, resPROD is residual of the regression of the logarithm of labour production on first nature variables.*

[11] This is the share of GDP density variance that is explained by gross second nature effect.
[12] This is the share of GDP density variance that is explained by net second nature effect.

In (): t-test

Table 6. Regression on first and second nature variables

Dependent variable: GDP per area						
Estimation	IV	OLS	OLS	OLS	IV	IV
Variables	*Italy*	*North*	*North East*	*North West*	*Centre*	*South*
Factor1	-.00008 (-0.05)	.0040*** (3.05)	.0031 (1.23)	.0041** (2.27)	-.00314 (-0.47)	.0046 (1.46)
Factor2	-.0179 (-1.38)	-.0210*** (-3.48)	-.0173* (-1.79)	-.0245*** (-2.93)	-.0873*** (-3.25)	-.0695** (-2.49)
Factor3	-.0158 (-0.71)	.0094 (0.97)	.0050 (0.31)	.0099 (0.69)	.04511 (0.98)	.0310 (0.89)
Res POP	.0152** (2.39)	-.0011** (-2.08)	-.0026*** (-2.83)	-.00091 (-1.03)	.0086** (2.15)	-.0093 (-1.59)
Res PROD	-.0393*** (-2.74)	.0016** (2.53)	.0033*** (2.94)	.0009 (0.83)	-.0190*** (-3.98)	-.0195* (-1.69)
Constant	1.1370*** (694.61)	1.0899*** (499.95)	1.1165*** (208.87)	1.0579*** (429.39)	1.1085*** (286.48)	1.1803*** (293.23)
R^2	0.3558	0.5620	0.6160	0.4574	0.4873	0.2901
Wald	2.55e+07***				1.19e+07***	9.14e+06***
F		3.90**	2.63**	2.03*		
Observations	1372	1165	610	555	256	650

****, **,*: significant at: 1%, 5%, 10resPOP is residual of the regression of the logarithm of population on first nature variables, resPRODxx resPOPxx is residual of the regression of the logarithm of labour production on first nature variables.*
In (): t-test

Our results show the relevance of first nature effects for the North of Italy, especially in the North-East. The second nature (population and GDP per worker) proves rather insignificant with only the South just over 1%. This result is not trivial and is supported by the unexplained variance that is very high and reaches 71% for the South of Italy; this circumstance may be due to the lack of relevant variables among the regressors. In ***Figures 3*** and ***4*** a summary of results by geographic area and effects is reported.

Figure 3. Analysis of Variance (ANOVA)

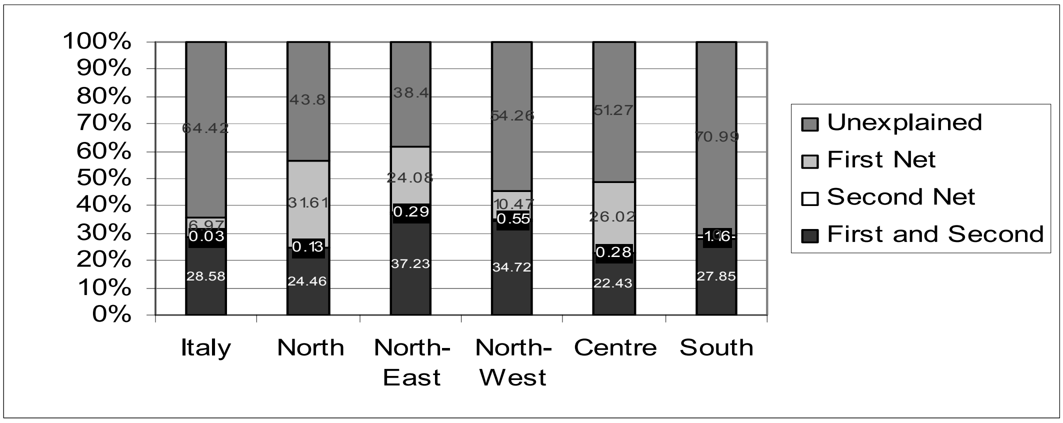
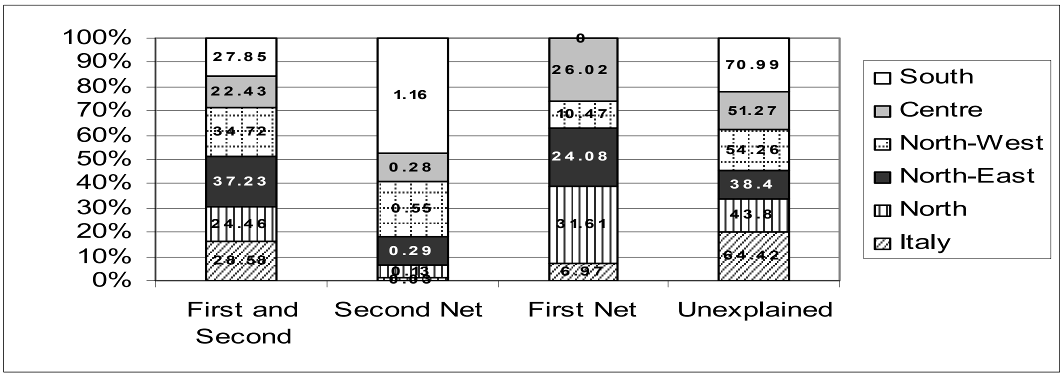


Figure 4. Significance of individual effects for the different macro-geographical areas



4. Concluding remarks

The aim of this paper was to ascertain the relevance of geographical features by focusing on agglomeration economies for the location of production of SLL in Italy. Hence our main concern was not to verify the relationship (or otherwise) between population and per capita income, but rather the impact of these variables in bringing about agglomeration processes. In other words, we investigated the causes of so-called agglomeration economies by using population and per capita GDP in order to control for the effects of second nature (or agglomeration economies).

With this intent, we applied a methodological approach based on Roos (2005), which proposes using **ANOVA** to infer the unobservable importance of first nature indirectly in a stepwise procedure. In order to disentangle first and second nature effects empirically, we controlled for second nature because every locational endowment is reinforced and overlaid by second nature advantages. Our results revealed that although there are major *gross second nature* effects, the *net second nature* effect appears to be relatively unimportant in determining the agglomeration process. However, what emerge as very important are *first nature effects* and also the *combined effect* of the two forces. Most of the GDP variance stemmed from unexplained factors, which is undoubtedly due to the lack of relevant variables. Of course, the fact that LLSs are fairly new on the scene and the lack of relevant observations at our disposal may well represent a limitation in our analysis.

In summary, we find that geographical features play a dominant role in leading to agglomeration processes, except in the South of Italy. This is to be expected, since we are not considering agglomerative processes related to the agricultural sector but rather to industry which is more concentrated in the North of the country. As regards the mixed effect of the interaction between first and second nature on GDP, we obtained a very high percentage (62% for North-East of Italy), which emphasizes that the two forces work well together to bring about agglomeration processes in LLSs.

References

Ades A.F, Glaeser, E.L., (1995), Trade and circuses: explaining urban giants. *Quarterly Journal of Economics* 110: 195–227.

Ayuda M.I., Collantes F, Pinilla V., (2005), From locational fundamentals to increasing returns: The spatial concentration of population in Spain, 1787-2000. *Documento de Trabajo 2005- 05, Facultad de Ciencias Económicas y Empresariales*, Universidad de Zaragoza (Spain).

Baum C, Stillman S, (2003), DMEXOGXT: Stata module to test consistency of OLS vs XT-IV estimates, *Statistical Software Components S401103*, Boston College Department of Economics.

Chasco C., Lopez A. M, Guillan R., (2008), The non-stationary influence of geography on the spatial agglomeration of production in the EU, *MPRA Paper No. 10737.*

Davidson R, MacKinnon J. G., (1993) Estimation and Inference in Econometrics, *New York: Oxford University Press.*

Faria J. R., Leon-Ledesma M., Sachsida A., (2006), Population and Income: Is There a Puzzle? *Journal of Development Studies*, Vol. 42, No. 6, 909–917 (August).

Funck R.H., (1995), Competition among locations: objectives, instruments, strategies, perspectives. In: Giersch H (ed.) Urban agglomeration and economic growth. Springer, *New York (USA) and Heidelberg (Germany), 227-255.*

Gallup J.L., Sachs J.D., (1998), Geography and economic development. Consulting Assistance on Economic Reform II *Discussion Paper 39, March*. http://www2.cid.harvard.edu/cidpapers/caer/paper39.pdf

Gallup J.L., Sachs J.D., Mellinger A.D., (1999), Geography and economic development. International Regional Science Review 22: 179-232.

Galor O., (2005), From Stagnation to Growth: Unified Growth Theory, *Handbook of Economic Growth, 171-293.*

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

Krugman P., (1993), First nature, second nature, and metropolitan location. *Journal of Regional Science* 33: 129-144.

Krugman P., (1999), The role of geography in development. *International Regional Science Review* 22: 142–161.

Limão N., Venables A.J., (2001), Infrastructure, geographical disadvantage, transport costs, and trade, *The World Bank Economic Review*, 15(3): 451-479.

Malthus, T. R. (1992) An Essay on the Principles of Population, original published in 1798 (Cambridge: Cambridge University Press).

Mathias K., (1980), Wirtschaftsgeographie des Saarlandes. *Buchverlag Saarbrücker Zeitung.*

McCallum J (1995) National borders matter: Canada–US regional trade patterns. *American Economic Review* 85: 615–623

Ottaviano G, J-F Thisse, (2004), New Economic Geography: what about the N?, *This research was supported by the Ministere de l'education, de la recherche et de la formation (Communaute française de Belgique), Convention 00/05-262*.

Rappaport, J., (2000), Is the speed of convergence constant? *Federal Reserve Bank of Kansas City Working Paper* 99-13 (August)

Rappaport J., Sachs J., (2003), The United States as a coastal nation. Journal of Economic Growth 8: 5-46

Roos M., (2005), How important is geography for agglomeration? *Journal of Economic Geography. Vol. 5 , No.5, pp.* 605-620.

Samuelson, P. A., (1988), Mathematical vindication of Ricardo on machinery, Journal of Political Economy, Vol. 96, 2, pp. 274–82.

Sargan J.D., (1964), Wages and prices in the United Kingdom: A study of econometric methodology. In Hart PE, Mill G, Whitaker JK (eds.) Econometric analysis for national economic planning, *Butterworths, London, UK.*

Appendix

variables		description
Second Nature	***l(gd)***	*logarithm of GDP per area*
	l(pop)	*logarithm of population*
	l(prod)	*logarithm of GDP per employed*
First Nature	***rain***	*average annual daily precipitation*
	agric	*this is a dummy variable that equals one if agricultural systems*
	metro	*this is a dummy variable that equals one if metro system*
	coast	*this is a dummy variable that equals one if coastal communities*
	tmax	*average annual maximum temperature*
	tmin	*average annual minimum temperature*
	sunsh	*average annual solar radiation*
	xcoo	*longitude earth coordinates*
	ycoo	*latitude earth coordinates*
	capreg	*this is a dummy variable that equals one if regional capitals*
	border	*this is a dummy variable that equals one if border regions*