A Mixed Geographically Weighted Approach to Decoupling and Rural Development in the EU-15

Francesco Pecci*, Maria Sassi**

*** Dipartimento di Economie, Società ed Istituzioni and SPERA – University of Verona – Italy**
**** Dipartimento di Ricerche Aziendali – University of Pavia – Italy**
e-mail: msassi@eco.unipv.it

Paper prepared for presentation at the 107th EAAE Seminar *"Modelling of Agricultural and Rural Development Policies"*. Sevilla, Spain, January 29th -February 1st, 2008

Abstract

The CAP reform and the recent EC communication aimed at preparing its Health Check emphasise the need for interventions locally based where agricultural policy integrates with a broader policy for rural areas growth. In this context, the paper investigates the possible different sets policy indicators affecting agricultural productivity at the regional level considering spatial heterogeneity by means of a Mixed Geographically Weighted Regression approach. The analysis is based on a set of policy sensitive indicators selected according to the key component of the CAP reform and referred to a sample of 164 EU-15 regions at NUTS2 level. The methodology adopted, new for the empirical literature on the topic, allows for a more accurate understanding of spatial relationship of the agricultural and socio-economic factors affecting agricultural productivity at the local level providing useful information for policy making..

Key words: CAP reform, agricultural productivity, spatial analysis, cluster analysis.

1. Introduction*

The reform of the first and second pillar of the Common Agricultural Policy (CAP) started in 2003 has emphasised the need for assessing its territorial impact and relationship with the other European policies, first of all the cohesion policy, and the Lisbon Strategy and Göteborg sustainability goals (European Commission 2004; 2005a; 2005b; 2005c; 2005d; 2005e).

The aims set out by the Commission of the European Communities, in the recent communication targeted at preparing the Health Check of the CAP reform, make the strengthening of the Rural Development policy necessary (Commission of the European Communities, 2007). This aspect further emphasises the need for interventions locally based where the agricultural policy is integrated with a broader development policy for the rural areas targeted at improving competitiveness for farming and forestry, environment and country side, and quality of life and diversification of the rural economy. The challenge for Member States' national rural development strategies becomes the identification of the areas where the use of the European support for rural development creates the most value added at the European Union (EU) level (Council Decision, 2006).

In this context, at least two issues relevant for the agricultural sector are emerging. They consist on the identification of a suitable set of policy sensitive indicators and on the understanding of the territorial dimension of their impact on the agricultural sector. Policy design in Member States requires explicit recognition of spatial heterogeneity in regional characteristics as well as in the heterogeneity of how these characteristics affect agricultural development. In this way, policy decisions can be spatially varied across regions for an effective local development.

The literature is conceptually aware of the problem but empirical analysis ignores or inadequately addresses the issue. This is particularly problematic for rural development analysis where the

* The research was supported by the Commission of the EU Genedec project (FP6-502184): *A quantitative and qualitative assessment of the socio-economic and environmental impacts of decoupling of direct payments on agricultural production, markets and land use in the EU.*

understanding of spatial heterogeneity of agricultural productivity marginal responses is desirable for policy decisions.

Standard approaches, such as Ordinary Least Squares or spatial econometrics, assume the marginal responses to explanatory variables fixed over space: there is one regression coefficient, a “global” parameter, for the entire sample. However, it can be expected that not only the explanatory variables ($x_{i,j}$) differ across space but that also the regression coefficients ($\beta_{i,j}$) are location specific. More precisely, variation in the total responses from a particular variable would be caused by variation in $x_{i,j}$, variation in $\beta_{i,j}$, and covariance between the two (Ali, Partridge, Olfert, 2007).

Concerning local variables, a further issue is of specific importance for rural development policy design even if still poorly addressed. Local variables might be spatial non-stationary: they have the same regression coefficients in sub-groups of generally neighbouring territorial units. Thus, it should be evaluated the possibility of networks across regions in policy design and implementation in order to reinforce actions through synergic effects. The aspect also contributes to the current debate on the definition of the concept of rural development areas and of their spatial borders.

In the light of these considerations, the paper provides a preliminary investigation of the possible sets of indicators affecting agricultural productivity at the regional level. More precisely, after the selection of a set of policy sensitive indicators according to key component of the reform of the CAP, focusing on a sample of 164 EU-15 regions at NUTS2 level, it:

- Identifies, by a Mixed Geographically Weighted Regression (MGWR) approach, the spatial non-stationary variables with an impact on agricultural productivity and the intensity of this impact; and
- Highlights, through a cluster analysis, the existence of groups of regions within which the level of agricultural productivity is affected by homogeneous values of the spatially non-stationary parameters.

The analysis is based on a previous papers prepared for the EU Genedec Project (FP6-502184) and of which it represents a methodological headway. The mentioned study is based on a Geographically Weighted Regression model where regression coefficients are all locally estimated. However, in practical cases some of the explanatory variables may be global in affecting agricultural development and only the remaining are local. The MGWR approach allows to distinguish between these two typologies of variables and in a second stage to underline within the local variables those that are spatially non stationary (Brunsdon, Fotheringham, Charlton, 1999). Thus, the methodology followed not only has never been adopted in the empirical literature on the topic, but it allows for a more accurate understanding of spatial relationship of the agricultural and socio-economic factors affecting agricultural productivity at the local level providing useful information for decision-makers.

2. Data Set

The selection of the indicators has taken into account the key components of the CAP reform of 2003 and 2004 and the reform of the Rural Development Policy for the programming period 2007-2013 in order to understand the agricultural and socio-economic policy sensitive variables. They make reference to the following areas: the EU agricultural support; agricultural innovation; agricultural efficiency and competitiveness; agricultural sustainability; economic development; structure of the labour market; infrastructure; territorial economic and social attraction capacity; and demographic features (Table 1).

Table 1. Indicators

Indicator	Source	Year	Indicator	Source	Year
Dependent variable			*TOTSUB*	Fadn	2000-2002
VALADD	Fadn	2000-2002	*COMPAY*	Fadn	2000-2002
Innovation			*SETPRE*	Fadn	2000-2002
- Research and Development			*SUBLIV*	Fadn	2000-2002
IPCAGR	Regio	2000-2002	**Economic development**		
KNOINT	Regio	2000-2002	*GDPIND*	Regio	2000-2002
MHTECH	Regio	2000-2002	**Labour market**		
- Human capital			*UNEMPR*	Regio	2004
LEARRU	Regio	2004	*EMPPER*	Regio	2004
EDUTER	Regio	2000-2002	*EMPRUR*	Regio	2002
Diversification			*SELFSH*	Regio	2004
INSEPA	Regio	2000-2002	*FEMALE*	Regio	2003
OTHGAI	Regio	2003	*PARTIME*	Regio	2004
Farm structure			**Infrastructure**		
HO3555	Regio	2003	*VEIPOP*	Regio	2000-2002
HO5005	Regio	2003	*BERUPO*	Eurostat	2004
BOVUAA	Regio	2000-2002	*PUBTOT*	Regio	2000-2002
CERULA	Regio	2000-2002	**Regional socio-economic attraction capacity**		
Environmental sustainability			*NETMIG*	Regio	2001-2003
SOIRIS	Jrc	2004	**Demographic features**		
WOODSL	Regio	2000-2002	*POPDEN*	Regio	2000-2002
EU intervention			*AGEING*	Regio	1998-2001

Important issues concerning the official data sources of reference, that is REGIO and FADN, need to be mentioned because they have strongly constrained the construction of the data set.

First, there is the lacking geographical breakdown. For this reason, at NUTS2 level important aspects cannot be quantified at all or even with a proxy. Among them there are agricultural production quality, capital and integration with the food chain; land and water quality; and infrastructures. In only few cases the constraint has been overcome making reference to national statistics due to the heterogeneous definition of the variables across EU Member States.

The issue has also affected the selection of the dependent variable. The agricultural productivity, in terms of agricultural working units, is not available for a large number of regions. Thus, the analysis has made reference to the farm net value added per utilised agricultural area (UAA).

The lacking geographical breakdown has had a further effect on the level of the regional articulation: some of the 164 regions of the sample have been taken at NUTS1 level. Even if their number is not large, this introduces a certain level of distortion in the analysis due to the different structure of the territorial units.

A final problem regards the unavailability of time series long enough for understanding the dynamic aspects of certain areas analysed, particularly those with structural characteristics. For this reason the analysis is static in the sense that it makes reference to a "central year", where indicators are average values for time periods included from 2000-2004, when possible, or values referred only to one year within that period.

2.1. The Agricultural Indicators

The agricultural indicators selected refer to innovation, efficiency, competitiveness, sustainability and the EU support within the CAP.

Research and Development (R&D) and human capital have the most significant impact on innovation. They are at the heart of the Lisbon Strategy, and thus understood as key contributors to the creation of a dynamic knowledge-based economy (Economic Commission, 2005f). The results from R&D should increase inputs productivity, support the introduction of new production methods and of improved institutional structures. On the other side, human resources are at the basis of the technological change. They depends strongly on the education level of workers and their life-long learning (Sassi, 2006a).

The innovation capacity of the agricultural sector has been approximated by the share of agricultural patents applications on total (*IPCAGR*). As innovation in agriculture is mostly imported from other sectors two indicators have been adopted in order to include the overall regional innovation capacity in the model. They are: the share of employment in total knowledge-intensive services on total employment (*KNOINT*) and of employment in high and medium high technology manufacturing sector on total employment (*MHTECH*).

Due to lack of data, the state and level of human capital in agriculture is difficult to fully comprehend. The aspect has been approximated by the state of life-long learning in rural areas represented by the share of 25-64 years hold participating in education and training (*LEARRU*). Also in this case, as for innovation, a specific variable has been introduced in order to take into account the level of education at the regional level: the share of students in the level 5 and 6 of education[1] on total students with less than 29 years old (*EDUTER*) has this function.

Diversification consists in the ability of farmers to have access to alternative sources of income (Sassi, 2006b). It has been approximated by two variables, the share of agricultural inseparable output on total agricultural output (*INSEPA*) and the share of farmers with other gainful activities on total (*OTHGAI*).

Farm structure underlines the efficiency and competitiveness of the farm sector, the well-being of farm households, the design of public policies and the nature of rural areas. It includes many dimensions among which farm organization, characteristics of farmers and their households, concentration of production, and tenure. Farm structure both affects and is influenced by policy interventions and economy at all levels.

The available data has allowed to consider the following variables in this area: the age structure in agriculture in terms of share of farmers less than 35 years old on those with more than 55 years old (*HO3555*), the physical farm size distribution ratio as share of farms with more than 50 ha of UAA on those with less than 5 (*HO5005*), the number of cows and beef on UAA (*BOVUAA*) and the ceral surfaces on UAA (*CERUAA*).

The age structure of farmers in combination with the importance of off-farm working provides preliminary information on the vitality and sustainability of the agricultural sector at the regional level (Vidal, Eiden, Hay, 2001).

Furthermore, *BOVUAA* and *CERUAA* can be understood as a proxy of the environmental sustainability of agriculture in the sense that they allow to emphasising crop and livestock intensity. However, in the area of environment, two specific variables have been introduced. They are the area at risk of soil erosion (Ton/ha/Year) (*SORIS*) and the woodland on total agricultural surface (*WOODSL*).

[1] According to the International Standard Classification of Education of 1997 level 5 is the first stage of tertiary education not leading directly to an advanced research qualification while level 6 is the second stage of tertiary education leading to an advanced research qualification (EUROSTAT, 2004).

Finally, the EU intervention has been considered through the share of total subsidies on UAA (*TOTSUB*) and its components, that is compensatory payments on UAA (*Compay*), livestock subsidies on UAA (*SUBLIV*), and set-aside premiums on UAA (*SETPRE*).

2.2 The Socio-Economic Indicators

The socio-economic context affecting agricultural productivity and relevant for decoupling and rural development has been taken into account considering the following areas: economic development, labour market, infrastructure, and territorial attraction capacity in terms of economic activities and population.

The level of economic development has been approximated by per capita GDP in PPS (*GDPIND*) that is the best estimate of the average regional income according to the available data.

Labour market has been represented in terms of rate of unemployment (*UNEMPR*), total employment (*EMPPER*), rural employment (*EMPRUR*), self-employment on total employment (*SELFSH*), part-time employment (*PARTIME*) and female unemployment (*FEMALE*) (OECD, 1996).

Infrastructure is another area where data is significantly lacking. Three proxies have been introduced: vehicles on total population (*VEIPOP*) as expression of the physical infrastructures; total bed places in hotels on total population (*BERUPO*) understood as tourist infrastructure; and employment in public sector on total employment (*PUBTOT*) considered as approximation of social infrastructures due to the fact that public sector also provides health, social care and education services.

The net migration ratio (*NETMIG*) shows the regional attraction capacity. The variable is linked with employment creation and quality of jobs, on the one side, and with quality of life, on the other (Bryden, Copus, MacLeod, 2002).

Finally, the demographic features have been represented by population density (*POPDEN*) and ageing index (*AGEING*) as measures of strengths and weaknesses of a region in the sense that a low level of population density and a high share of elderly people can be interpreted as a signal of the fragility of an area and vice versa.

3. Methodology

3.1 Mixed Geographically Weighted Regression

Geographically weighted regression (GWR) is a useful technique to explore spatial nonstationarity (Fotheringham et al, 2002) by calibrating a varying coefficient regression model with the form

(1) $$y_i = a_0(u_i, v_i)\sum_{j=1}^{p}\beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where y_i are the observed dependent variables, (x_{i1}, x_{i1},…, x_{ip}) the explanatory variables at the location (u_i,v_i) in the studied area and ε_i are the error terms that are assumed to be independent and normally distributed with zero mean and common variance σ^2.

Considering the situation where some explanatory variables influencing the response may be global, while others are local, Brundson et al. (1999) have proposed a model, called mixed GWR (MGWR), in which some coefficients are assumed to be fixed, the others are allowed to vary across the regions. An MGWR model is in the form

(2) $$y_i = \sum_{j=1}^{q} \beta_j x_{ij} \sum_{j=q+1}^{p} \beta_j (u_i, v_i) x_{ij} + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

setting $x_{i1} = 1$ or $x_{i,q+1} = 1$, the intercept is a constant or spatially varying.

The calibration of a MGWR model, as proposed in Fotheringham et al. (2002) is summarized below in matrix notation

$$\mathbf{X}_G = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix}, \quad \mathbf{X}_L = \begin{pmatrix} x_{1,q+1} & x_{1,q+2} & \cdots & x_{1p} \\ x_{2,q+1} & x_{2,q+2} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,q+1} & x_{n,q+2} & \cdots & x_{np} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\boldsymbol{\beta}_G = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \quad \boldsymbol{\beta}_L(u_i, v_i) = \begin{pmatrix} \beta_{q+1}(u_i, v_i) \\ \beta_{q+2}(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{pmatrix}, \; i = 1, 2, \ldots, n,$$

and

(3) $$\mathbf{S}_G = \mathbf{X}_G (\mathbf{X}_G^T \mathbf{X}_G)^{-1} \mathbf{X}_G^T, \qquad \mathbf{S}_L = \begin{pmatrix} \mathbf{x}_{L1}^T \left[\mathbf{X}_L^T \mathbf{W}(u_1, v_1) \mathbf{X}_L \right]^{-1} \mathbf{X}_L^T \mathbf{W}(u_1, v_1) \\ \mathbf{x}_{L2}^T \left[\mathbf{X}_L^T \mathbf{W}(u_2, v_2) \mathbf{X}_L \right]^{-1} \mathbf{X}_L^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}_{Ln}^T \left[\mathbf{X}_L^T \mathbf{W}(u_n, v_n) \mathbf{X}_L \right]^{-1} \mathbf{X}_L^T \mathbf{W}(u_n, v_n) \end{pmatrix}$$

where $\mathbf{x}_{Vi}^T = (x_{i,q+1}, x_{i,q+2}, \ldots, x_{ip})$ is the ith row of $\mathbf{X}_L$ and

(4) $\mathbf{W}(u_i,v_i) = \mathrm{diag}[w_1(u_i,v_i), w_2(u_i,v_i), \ldots, w_n(u_i,v_i)]$

is an $n \times n$ diagonal weight matrix at location (u_i,v_i) (u_i,v_i are the geographic coordinates of each region), and the weights are taken as a function of the distance from (u_i,v_i) to other analysed regions. The element of the weight matrix are calculated with a bi-square function (Fotheringam et al., 2002)

(5) $w_{ij} = [1-(d_{ij}/b)^2]^2$ if $d_{ij} < b$

$= 0$ otherwise

where b is referred to as the bandwidth. If i and j coincide, the weighting of data at that point is equal to unity and the weighting of other data decrease according to a Gaussian curve as the distance between i and j increases. An exhaustive discussion of the matrix $\mathbf{S}_L$ is in Leung et al. (2000).

The procedure to calibrate a MGWR model, as proposed by Fotheringam et al. (2002), produces the estimates of the constant coefficient vector $\hat{\boldsymbol{\beta}}_G$ as

$$\hat{\boldsymbol{\beta}}_G = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_q)^T = [\mathbf{X}_G^T (\mathbf{I} - \mathbf{S}_L)^T (\mathbf{I} - \mathbf{S}_L) \mathbf{X}_G]^{-1} \mathbf{X}_G^T (\mathbf{I} - \mathbf{S}_L)^T (\mathbf{I} - \mathbf{S}_L) \mathbf{Y}$$

and the spatially varying coefficient vector at location (u_i,v_i) as

$$\hat{\boldsymbol{\beta}}_L(u_i, v_i) = [\beta_{q+1}(u_i, v_i), \beta_{q+2}(u_i, v_i), \ldots, \beta_p(u_i, v_i)]$$
$$= [\mathbf{X}_L^T \mathbf{W}(u_1, v_1) \mathbf{X}_L]^{-1} \mathbf{X}_L^T \mathbf{W}(u_1, v_1)(\mathbf{Y} - \mathbf{X}_G \hat{\boldsymbol{\beta}}_G)], \quad i = 1, 2, \ldots, n.$$

Finally, the fitted values at n location are

(6) $$\hat{\mathbf{Y}} = (\hat{y}_{1,}\hat{y}_{2,}...\hat{y}_n)^T = \mathbf{S}_L(\mathbf{Y} - \mathbf{X}_G\hat{\boldsymbol{\beta}}_G) + \mathbf{X}_G\hat{\boldsymbol{\beta}}_G = \mathbf{S}_L\mathbf{Y} + (\mathbf{I} - \mathbf{S}_L)\mathbf{X}_G\hat{\boldsymbol{\beta}}_G$$

3.2. Cluster Methodology

Data mining computerized methods based on cluster analysis have been followed in the study in order to classify regions according to an homogeneous profile in terms of marginal responses of agricultural productivity to the explanatory variables. This methodology identifies groups of statistic units characterised by internal cohesion and external distance, it is, maximizing both the internal cluster homogeneity and the inter-cluster heterogeneity.

According to the literature, the analysis has been articulated into three steps: model specification, comparison and interpretation.

For the specification of the model two non hierarchical cluster approaches have been compared: the k-means algorithm for a number of clusters equal to six and a 2x3 Kohonen map. In order to prevent the results from being influenced by the units of measurement of the indicators, by giving a major weight to the highest distances, the variables have been standardised.

The two models have been compared by splitting the total variability into within-group variability and between-group variability, leading to the overall R^2 and to the R^2 for the specific parameters object of classification. The comparison has favoured the Kohonen Maps. This latter seems to be a better choice also from an economic point of view. The algorithm selected has the advantage to define more distinct groups determined by a distinct behaviour than those from k-means clustering that are due to randomness.

A Kohonen network is formed by two levels of neurons: a first one of incoming neurons and a second and bi-dimensional one (Kohonen, 1997, Kohonen et al., 1994). The incoming level is used to calculate the total weight of the input, whereas the bi-dimensional one calculates the output of the net.

Considering $w_{ij(t)}$ as the weight between the input for the neuron in the i position and the output of the neuron where

$0 \leq i \geq n - 1$

n = number of input

t = step in the learning model

if $N_{i(t)}$ is the number of neurons close to the j position and if $x_{i(t)}$ is the input in the i position, the learning algorithm is as follows:

a. the map dimensions are defined by establishing the weights $w_{ij(t)}$ between 0 and 1 initially and fixing the value of $N_{i\,(0)}$ as high as possible;

b. presentation of an input $x_{0(t)},\ x_{1(t)},\ x_{2(t)},\ \ldots,\ x_{n(t)}$ for which its values multiplied by the respective synaptic weight represents the stimulus given to the neuron in the network of Kohonen;

c. the Euclidian distances are calculated, $d2j$, between the input and each neuron of output j;

d. the successful neuron, $j*$, is selected. It is, the one matching the minimum distance or the higher activation value;

e. the weights are modified from the neuron of input to the $j*$ neuron and to those close to it2 defined into the $N*_{i\,(t)}$. The new weights are given by

---

[2] The fact that even the neurons being close to j* have been modified, derives from the network's property to generalize. In fact, the network tries to create regions constituted by a large amount of values that lie around the input. In this way, the vectors being closer to the training values are properly classified. This concept is not present in the traditional classification methodologies.

$w_{ij(t+1)} = w_{ij(t)} + n_{(t)} [x_{i(t)} - w_{ij(t)}]$

where $n_{(t)}$, which is smaller than 1 and higher than 0, is the velocity of adjustment. It decreases over time in order to progressively decrease weights adjustment.

f. back to step b. (Giudici, 2004).

Because of the existence of vicinity and the sensitivity to history of this algorithm the result is a homogeneous classification of the observations rarely characterised by relatively large groups coexisting with relatively small ones. The used learning algorithm depends on the frequency of past allocation allowing to solve the problems of the elephant cluster, i.e. an over dimensioned class in terms of relative number of observations.

A SOM works by smoothing the seeds in a manner similar to kernel estimation methods, but smoothing is done in neighbourhoods in the grid space rather than in the input space (Mulier, Cherkassky, 1995).

Finally, the number of clusters has been firstly decided applying to the Ward method and to the statistic R^2 and then evaluating the result in the light of economic considerations.

4. Results

4.1. MGWR results

To identify potentially significant variables, GWR regression were performed first to test the relationship between the dependent variable and each of the independent variables. The kernel bandwidth, adaptive in our case, is determined by Akaike Information Criterion (AIC) minimization. The variables were sorted in ascending order on the basis of the AIC values. Contemporarily, we have performed a Monte Carlo nonstationary significance test for the single variable GWR models to verify the spatial stationarity of variables. The variable that do not show significant spatial variation are global variables, while the variables that significantly varying across the space are local variables. To detect the presence of multicollinearity, we have calculated in advance the variance inflation factor (VIF), as a OLS model, and the variables with the more elevated values of VIF have been abandoned.

The MGWR model estimate is the following:

$$VALADD_i = b_0(i) + b_{g1}POPDENi + b_{g2}MHTECH_i + b_{g3}CERUAA_i + b_{g4}EDUTER_i + b_{g5}BERUPO_i + b_{g6}GDPIND_i + b_{l1}(i)WOODSL_i + b_{l2}(i)VEIPOP_i + b_{l3}(i)IPCAGR_i + b_{l4}(i)INSEPA_i + {}_lb_5(i)HO5005_i + b_{l6}(i)UNEMPR_i + b_{l7}(i)SOIRIS_i + b_{l8}(i)NETMIG_i + b_{l9}(i)HO3555_i + b_{l10}(i)OTHGAI_i + b_{l11}(i)PUBTOT_i + b_{l12}(i)BOVUAA_i + b_{l13}(i)TOTSUB_i$$

where b_0 are the intercept terms varying across the regions, $b_{g(1\ to\ 6)}$ are the global parameters of the independent variables not varying across the space; $b_{l(1\ to\ 13)}(i)$are the local parameters of the independent variables varying across the space. The results are illustrated in Table 2.

According Brundson et al. (1999), we compare the range of values of the local estimates between the lower and upper quartiles with the range of values at ±1 standard deviations of the respective global estimate , that correspond to 2 × S.E. of each global estimate (Table 3).

If the range of local estimates between the inter-quartile range is greater than that of 2 standard errors of the global mean, this suggests the relationship might be non-stationary.

For *INTCPT*, *WOODSL*, *VEIPOP*, *HO3555*, *SOIRIS*, *OTHGAI*, *PUBTOT* and *TOTSUB* the interquartile range of the local estimates is much greater than 2 × S.E. indicating a non-stationary relationship.

Table 2. Parameters of EU-15 MGWR model

Variable	*Min.*	*Lwr Quart.*	*Median*	*Upr Quart.*	*Max.*	*Global*
INTCPT	-0.9020	-0.0637	0.1724	0.4006	0.9261	-
POPDEN	-	-	-	-	-	0.1476
MHTECH	-	-	-	-	-	-0.0067
CERUAA	-	-	-	-	-	-0.1791
EDUTER	-	-	-	-	-	-0.0688
BERUPO	-	-	-	-	-	-0.0556
GDPIND						-0.0167
WOODSL	-2.4732	-0.2298	-0.0240	0.0903	0.6087	-
VEIPOP	-1.7767	-1.4038	-0.2566	-0.0614	0.2093	-
IPCAGR	-0.0221	0.0806	0.1534	0.3058	0.7656	-
INSEPA	-1.5498	-0.1502	-0.0589	0.0478	0.6055	-
HO5005	-1.7679	-0.8013	-0.6101	-0.4171	-0.0102	-
UNEMPR	-1.1722	-0.2795	-0.1581	-0.0745	0.6427	-
SOIRIS	-1.6440	-0.4785	-0.2465	0.0770	1.5656	-
NETMIG	-0.2588	0.0656	0.1694	0.3916	0.6600	-
HO3555	-0.7130	-0.3057	-0.1938	0.0239	0.3313	-
OTHGAI	-0.6688	-0.4722	-0.2812	-0.1429	0.0736	-
PUBTOT	-0.5192	-0.2635	-0.0656	0.3121	0.8228	-
BOVUAA	-0.2074	-0.0782	0.0136	0.1844	0.7629	
TOTSUB	-0.6077	-0.0469	0.1292	0.3093	0.9657	

(*)Spatial variability test of local variables: ***, signif.; °°° not signif.

Table 3. Spatial variability of the local variables

Variable	Int. quant. range	2 × S.E.	SP-test
INTCPT	0.4643	0.1022	***
WOODSL	0.3201	0.1242	***
VEIPOP	1.3424	0.1102	**
IPCAGR	0.2252	0.1191	n/s
INSEPA	0.1980	0.1242	*
HO5005	0.3842	0.1228	**
UNEMPR	0.2050	0.1282	*
SOIRIS	0.5555	0.1153	***
NETMIG	0.3259	0.1316	n/s
HO3555	0.3296	0.1411	n/s
OTHGAI	0.3293	0.1199	*
PUBTOT	0.5756	0.1354	***
BOVUAA	0.2626	0.1221	n/s
TOTSUB	0.3562	0.1246	***

*** signif. at .1%; ** signif. at 1%; * signif. at 5%

Successively, we examine the significance of the spatial variability in the local parameter estimates more formally by a Monte Carlo test.

The results of this test (col. 3 Table 3) on the local estimates indicates that there is significant spatial variation in the local parameter estimates for the preceding variables to which are added *INSEPA* and *UNEMPR*. These results reinforce the conclusions reached above with the informal examination of local parameter variation and this shows like significant socio-economic variables that explain the variability of the agriculture value added per hectare have remarkable local characteristics in EU 15.

The spatial variation in the remaining variables is not significant and in each case there is a reasonably high probability that the variation occurred by chance.

4.2. Testing for collinearity

The GWR approach for the local variables shows a significant improvement, in term of residual sum of square (RSS), 12.5, with respect to ordinary least square (OLS), 64.2, while the F value of the ANOVA test value, as proposed by Fortheringham, Brunsdon, Charlton (2002), is 5.34 (p-value = 0.000).
When using GWR approach it is possible calculate VIF values, as a collinearity diagnostic, for each explanatory variable for each local regression. The VIF for a variable *v* at location *i* is

(7) $$VIF_v(i) = \frac{1}{1 - R_v^2(i)}$$

where $R_v^2(i)$ is the coefficient of determination when x_v is regressed on the other explanatory variables at location *i*. In our case the mean values of the VIF (Table 4) of the local variables are not acceptable.

Table 4. Mean of VIF values with GWR approach

HO5005	*IPCAGR*	*TOTSUB*	*HO3555*	*OTHGAI*	*BOVUAA*	*WOODSL*
4.935	2.746	5.289	6.274	4.112	3.189	4.2811
UNEMPR	*INSEPA*	*SOIRIS*	*NETMIG*	*PUBTOT*	*VEIPOP*	
4.752	5.0261	5.803	4.547	5.102	3.969	

VIF don't consider collinearity with the constant term and don't clarify the nature of the collinearity and in the GWR approach with more two explanatory variables is very difficult to interpret the VIF values. Belsley (1991) suggest another diagnostic tool for collinearity that uses SVD of the design matrix **X, X** = **UDV**^{T}, where **U** contains the eigenvectors of **X** and **D** is a diagonal matrix containing eigenvalues, to form condition indexes of this matrix and variance-decomposition proportions of the coefficient covariance matrix. The diagnostic is capable of determining the number of near linear dependencies in the data matrix **X**, and the diagnostic identifies which variables are involved in each linear dependency. For diagnostic purposes the singular value decomposition is applied to the variance-covariance matrix of the least-squares estimates and rearranged to form a table of variance-decomposition proportions. Belsley outlines that a large value of the condition index is associated with each near linear dependency, and the variables involved in the dependency are those with large proportions of their variance associated with large condition indexes; the variance-decomposition proportions in excess of 0.5 indicate the variables involved in specific linear dependencies. The joint condition of condition index > 30 and variance-decomposition proportions > 0.5 diagnose the presence of strong collinear relations as well as determining the variables involved. In the GWR framework SVD of design matrix is (Wheeler, 2007)

(8) $\mathbf{W}^{½}(i)\, \mathbf{X} = \mathbf{UDV}^T$

where $\mathbf{W}^{½}(i)$ is the square root of the diagonal weight matrix at location *i* calculated from the kernel function.

Table 5 shows the condition indexes and variance-decompositions proportions for the largest variance component for the observation with a condition index greater than 15, only for the variables with variance-decomposition proportions that exceeds 0.5. The joint conditions of condition index > 15 and variance-decompositions proportions > 0.5 indicate that collinearity doesn't disturb our model (only the variance-decomposition proportions for PUBTOT shows values > 0.5 for the Greek Regions).

Table 5 - Condition indexes > 15 and variables with variance-decomposition proportion > 0.5 (bold)

NUTS2	*Condition Index*	*HO3555*	*PUBTOT*
DE73	15.383	0.046	0.101
DE80	15.300	**0.828**	0.136
GR11	16.961	0.103	**0.734**
GR12	15.835	0.113	**0.759**
GR13	15.481	0.121	**0.765**
GR14	16.115	0.115	**0.751**
GR23	15.615	0.108	**0.743**
GR24	16.156	0.109	**0.745**
GR25	16.170	0.107	**0.734**
GR30	17.268	0.106	**0.729**
GR41	17.583	0.092	**0.724**
GR42	19.516	0.091	**0.695**
GR43	17.734	0.102	**0.708**

A further analysis is to explore the overall correlation between the set of local regression coefficient; the absolute value of these correlations greater than 0.65[3]: are: *INTCPT* versus *WOODSL* (0.703), *OTHGAI* versus *VEIPOP* (0.685) and *PUBTOT* versus *VEIPOP* (0.799). Figure 1.a shows the scatter plots for *PUBTOT* versus *VEIPOP*. Then, as suggest by Wheeler and Tiefelsdorf (2005), we have calculated the local coefficient correlations to verify the existence of outlying clusters that are mapped in Figure 1.b.

The joint analysis of the scatter plots and choropleth map shows that for *PUBTOT-VEIPOP* support that in some areas the absolute magnitude of local coefficient correlation is greater than 0.75, above all the Greek Regions for *PUBTOT-VEIPOP*. That will be considered during the analysis of the implication of the MGWR answers.

4.3. Classification result

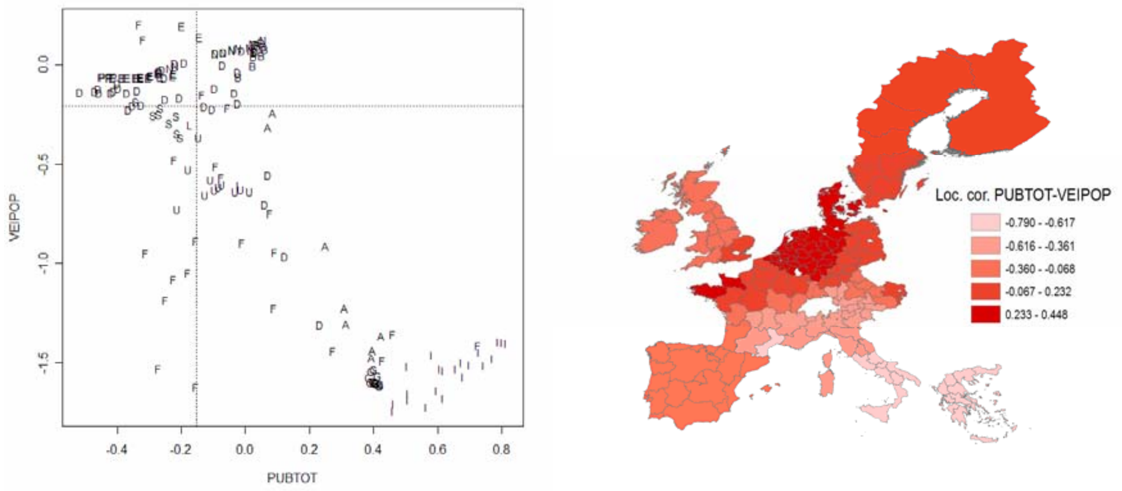
The Ward method and the R^2 statistics have suggested six clusters as optimal. The number of cases in each cluster is shown in Figure 2 and underlines a quite homogeneous number of observations in each cluster.

A part from two indicators, *OTHGAI* and *TOTSUB*, all the others have resulted important to the formation of the cluster even if with a different intensity[4]. This means that the regional impact of these parameters on the agricultural productivity is combined with their spatial proximity.

---

[3] The Regions are represented with the word-initial of the Country (Finland =L).

[4] A decision tree calculates the relative importance values that can assume values between 0 (no contribution to the cluster profile) and 1 (maximum contribution to the formation of the cluster). The intensity of the importance of the indicators is the following: *INTERCEPT* (0.7219), *INSEPA* (0.1856), *OTHGAI* (0), *HO5005* (0.6416), *SORIS* (0.7456), *WOODSL* (0.2562), *TOTSUB* (0), *UNEMPR* (0.8719), *VEIPOP* (0.1846), *PUBTOT* (1.0000).

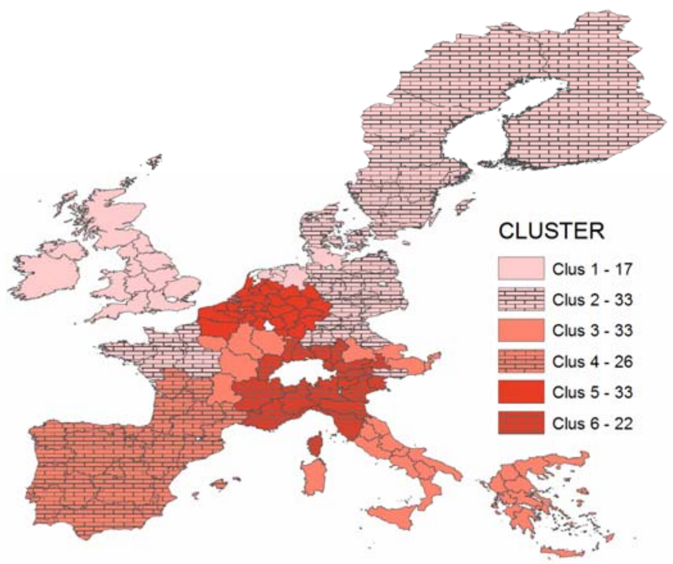


The dotted lines are the levels of the relate global parameter estimates

1.a . Scatter plot (r = -0.799)

1.b. Local coefficient correlation

Figure 1. Local estimated regression coefficients *PUBTOT* and *VEIPOP*



Cluster 1: DE94, DEF0, IE00, NL11, NL12, NL13, UKC0, UKD0, UKE0, UKF0, UKG0, UKH0, UKJ0, UKK0, UKL0, UKM0, UKN0;
Cluster 2: AT21, DE11, DE12, DE22, DE23, DE24, DE25, DE26, DE40, DE80, DE91, DE92, DE93, DED0, DEE1, DEE2, DEE3, DEG0, DK00, FI00, FR23, FR24, FR25, FR51, FR52, SE0A, SE01, SE02, SE04, SE06, SE07, SE08, SE09 ;
Cluster 3: AT11, AT22, AT31, DE21, FR10, FR21, FR26, FR41, FR72, GR11, GR12, GR13, GR14, GR21, GR22, GR23, GR24, GR25, GR30, ,GR41, GR42, GR43, ITE2, ITE3, ITE4, ITF1, ITF2, ITF3, ITF4, ITF5, ITF6, ITG1, ITG2;
Cluster 4: ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, FR53, FR61, FR62, FR63, FR81, PT11, PT15, PT16, PT17, PT18;
Cluster 5: BE21, BE22, BE23, BE24, BE25, BE31, BE32, BE33, BE34, BE35, DE71, DE72,
DE73, DEA1, DEA2, DEA3, DEA4, DEA5, DEB1, DEB2, DEB3, DEC0, FR22, FR30, NL21, NL22, NL23, NL31, NL32, NL33, NL34, NL41, NL42;
Cluster 6: AT21, AT32, AT33, AT34, DE13, DE14, DE27, FR42, FR43, FR71, FR82, FR83, ITC1, ITC2, ITC3, ITC4, ITD1, ITD2, ITD3, ITD4, ITD5, ITE1.

Figure 2. Cartographic presentation of the classification result

In this context, certain parameters play a major role in the final regionalization results showing a relative importance value greater than 50% . They are related to social infrastructures (*PUBTOT*), labour market (*UNEMPR*), environmental sustainability (*SOIRIS*), and farm structure (*HO5005*).
The interpretation of the results is based on the cluster profiles pointed out by the analysis and their spatial representation. More precisely, each sub-group of regions has been represented in a map with a different colour (Figure 2) and the profile has been characterised comparing the input mean for each cluster to the overall means emphasising the variables whose parameters are greater than the overall means. The latter can be understood as interesting policy sensitive areas for the agricultural development not only at the regional level but also at the level of a specific sub-group of regions.
The agricultural value added results strongly sensitive to:
- Physical infrastructures (*VEIPOP*) in Cluster 1;
- Farm structure (*HO5005*), environmental sustainability (*SOIRIS*) and physical infrastructures (*VEIPOP*) in Cluster 2;
- Agricultural diversification (*INSEPA*, *OTHGAI*) and sustainability (*WOODLS*) and social infrastructure (*PUBTOT*) in Cluster 3;
- Farmers diversification in other gainful activities (*OTHGAI*), unemployment (*UNEMPR*), physical infrastructure (*VEIPOP*) and total subsidies (*TOTSUB*) in Cluster 4;
- Diversification in inseparable activity (*INSEPA*), agricultural sustainability (*WOODLS*) and physical infrastructures (*VEIPOP*) in Cluster 5;
- Diversification (*INSEPA*, *OTHGAI*), agricultural sustainability (*WOODLS*) and social infrastructure (*PUBTOT*) in Cluster 6.

The map in Figure 2 underlines distinct regional and spatial coherence although a great diversity between clusters in terms of marginal responses of the agricultural productivity to the explanatory variables considered. The aspect suggests the operational of specific characteristics that seems to be linked to the national and sub-national level. In part, the result should depend on the fact that some of the variables selected reflect historical, physical and spatial conditions that are strongly territorial related. These factors to become strategic components of a broad strategy of agricultural and rural development need to be activated by interventions calibrated at the local level. This consideration has a particular meaning in the case of clusters defined at the sub-national level. Agricultural productivity might be affected differently across regions in a single Member State if policy design is referred to the national level. Thus, the importance to consider regional and sub-regional sometime large diversity of situations in the National Strategic Plans finds confirmation.
A further consideration refers to the intercept resulted locally non-stationary suggesting the operational of other variables strongly dependent to spatial conditions that however are difficult to be quantified according to the available data.

5. Conclusions

The analysis developed points out interesting aspects that can contribute to the current debate on the new challenges form the Health Check of the CAP reform. The methodology adopted, based on the MGWR approach, has allowed to distinguish the variables with a global and a local impact on agricultural productivity and, within the latter, those spatial non-stationary more accurately than what enabled by the GWR approach. It also suggests that regional policy analysis that have ignored the spatial heterogeneity of the marginal responses of the agricultural productivity to the explanatory variables have mispresented the actual patterns.

In this context, the EU direct support, even if a local variable spatially non-stationary, has resulted a relevant measure for agricultural productivity at the territorial level only in a restricted number of regions mainly concentrated in Spain, Portugal and part of France. Adding to this, the territorial distribution of the resources of the first pillar unbalanced towards the northern regions, the suitability of direct payments not only as instrument for rural development, but also for achieving the other EU targets, particularly cohesion, is questionable.

On the contrary, results have suggested the potential importance of the Rural development policy for agricultural growth. As expected, agricultural value added at the regional level is affected by policy variables of both agricultural and socio-economic nature. Among them, agricultural innovation and diversification impact significantly on the sector productivity in a wide number of regions even if with a different intensity.

These aspects suggest the possible role that modulation, regionalization and multi-sector and multi-region interventions can play in the overall framework of the CAP in order to face the problem of the territorial distribution of the EU funds. The Health Check might represents an important appointment for making this role effective through a better specification and reinforcement of modulation and regionalization and a stronger integration of the CAP with the other policies. At the same time Member States and Regions should strengthen their policy design capacity and political will, the preconditions and indispensable components for supporting interventions aimed at realised a locally based bottom-up approach to rural development. In this context, the analysis has also underlined the potential importance of establishing networks among Member States and Regions. The aspect has already been emphasised within the reform of the Rural Development policy (European Commission, 2005e). However, according to the results achieved networks should go further the implementation, evaluation and exchange of best practices. It should be evaluated the suitability to extend them to the phase of policy design.

These observations has a specific meaning in the light of the next process of the EU budget review and the possible further cut in the expected Rural development support and require a careful refocus of support from the new rural development fund on growth, jobs and sustainability.

A final remark concerns the modelling approaches and tools for impact assessment of agricultural and rural development policies. As previously underlined, the results have pointed out the importance of a multi-sector and multi-region approach to agricultural and rural development that suggest the need for a better interaction between partial equilibrium and regional computed general equilibrium models not only for policy analysis of the direct and indirect effects that different options of public interventions should have on the socio-economic context, but also for consensus building.

REFERENCES

Ali, K., Partridge, M.D. and Olfert, M.R. (2007). Can Geographically Weighted Regressions Improve Regional Analysis and Policy Making?. *International Regional Science* 30, 3: 300-329.

Belsley, D.A. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression.* New York: John Wiley.

Brunsdon, C., Fotheringham, A.S. and Charlton M.E. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science* 39: 497-524.

Bryden, J.M., Copus, A. and MacLeod, M. (2002). Rural Development Indicators, Report for Eurostat with LANDSIS. Luxembourg, February.

Commission of the European Communities (2007). Communication from the Commission to the Council and the European Parliament. Preparing for the "Health Check" of the CAP reform. COM(2007)722.

Council Decision (2006). Community Strategic Guidelines for Rural Development (Programming Period 2007 to 2013). 20 February.

European Commission (2002b). Productivity: the Key to Competitiveness of European Economies and Enterprises. COM(2002)262 final, Brussels.

European Commission (2004). The Common Agricultural Policy Explained. http://europa.eu.int.

European Commission (2005a). Agriculture – CAP reform. http://europa.eu.int.

European Commission (2005b). Lisbon Strategy – Putting Rural Development to Work for Jobs and Growth. http://europa.eu.int.

European Commission (2005c). The Common Agricultural Policy and the Lisbon Strategy. http://europa.eu.int.

European Commission (2005d). Lisbon Strategy. http://europa.eu.int.

European Commission (2005e). Cap Reform – Strategic Guidelines for Rural Development. http://europa.eu.int.

European Commission (2005f). Lisbon Strategy – Putting Rural Development to Work for Jobs and Growth. http://europa.eu.int.

Eurostat (2006). European Regional and Urban Statistics. Reference Guide. Luxembourg: European Communities.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships.* West Sussex: Wiley.

Giudici, P. (2004). *Data Mining.* West Sussex: Wiley.

Kohonen, T. (1997). *Self-Organizing Maps.* Springer-Verlag: Berlin

Kohonen, T., Makisara, K. and Saramaki, T. (1994). Phonotopic Maps-insightful Representation of Phonological Features for Speech Recognition. Preecidings of 71CPR, International Conference on Pattern Recognition Los Alamitos, CA, IEEE Computer Soc. Press.

Leung, Y., Mei, C.L. and Zhang, W.X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning* A32: 9-32.

Mei, C.L., Wang, N. and Zhang, W.X. (2006).Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning* A 38: 587-598.

Mulier, F. and Cherkassky, V. (1995). Self-Orgation as an Interative Kernel Smoothing Process. *Neural Computation* 7: 1165 - 1177.

OECD (1996). *Territorial Indicators of Employment: Focusing on Rural Development.* Paris: OECD.

Sassi, M. (2006a). Agricultural convergence and competitiveness in the eu-15 regions. paper presented at the International Conference IAAE2006, Gold Coast (Australia).

Sassi, M. (2006b). Has Competitiveness Supported Agricultural Cohesion Across the EU-15 Regions? The Last Decade Experience. paper presented at the USDA and AIEA2 International Meeting on Competitiveness in Agriculture and the Food Industry: US and EU Perspectives, Bologna.

Vidal, C., Eiden, G. and Hay, K. (2001). Agriculture as a Key Issue for Rural Development in the European Union. http://www.unece.org/stats/documents/ces/ac.61/2001/wp.2.e.pdf

Wheeler, D.C. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning* A 39: 2464-2481.

Wheeler, D.C. and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geograpical Systems* 7: 161-187.