Looking for Regional Convergence: Evidence from the Italian Case with Multivariate Adaptive Regression Splines

Iacopo Odoardi[1(✉)], Fabrizio Muratore[2], Edgardo Bucciarelli[1], and Shu-Heng Chen[3]

[1] Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, Section of Economics and Quantitative Methods, University of Chieti-Pescara, Pescara, Italy
{iacopo.odoardi,edgardo.bucciarelli}@unich.it
[2] National Foundation of Accountants - Economics and Statistical Office, Rome, Italy
muratore@fncommercialisti.it
[3] Department of Economics, National Chengchi University, Taipei, Taiwan
chchen@nccu.edu.tw

Abstract. This paper examines the role of data mining analysis in explaining the Italian regional dualism with the aim of suggesting economic policies to fill the existing socio-economic gaps. We analyze the 2004–2014 period exploiting the capacity of MARS model in finding relationships among data. In Italy, the presence of a North-South divide is well-known for decades and present for several social and economic aspects. Recent studies prove that strong differences exist also in the regional human capital. Thus, we search for the causes of the local differences, also considering the entrepreneurial vitality and the international trade leverage. Among several variables, MARS is useful in showing the actual determinants on which to intervene. This is possible by comparing regions grouped homogeneously into clusters using recent data. MARS results are used for policy suggestions with the aim of filling the income gap.

Keywords: Regional convergence · Clusters · MARS · Regional policy

1 Introduction and the Italian North-South Dualism

The North-South divide in Italy is a problem studied for decades [10], strongly increased after the Italian unification [9]. Since that period, there was a progressive distancing of the economic development paths of the Northern regions with respect to the South. The North has been able to encourage industrial settlements, initially thanks to the natural resources, afterwards by supporting aggregate demand also thanks to the geographical proximity to European and international markets [1]. For Iuzzolino *et al.* [23] the last observed convergence period among the Italian regions was in the twenty years after the Second World War, while for Barro and Sala-I-Martin [5], in Italy and in other developed countries, there are some traces of convergence until the 1980s. Therefore, two different contexts must be analyzed in Italy, in which the differences between the population of the North and South also exist in the trust and cooperative behavior [8] and in the cultural and educational background [24].

E. Bucciarelli et al. (eds.), *Decision Economics*, Advances in Intelligent Systems and Computing 618, DOI 10.1007/978-3-319-60882-2_10

A specific data mining tool can improve the economic analysis. We use a MARS model (multivariate adaptive regression splines, [18]) to observe the nonlinear and multidimensional relationship between the income and several regressors at the regional level. We found MARS more efficient than traditional techniques. It is used in many fields of research in decision-making and forecasting models [25]. For example, Abraham *et al.* [2] argue that MARS is more efficient compared to several recent models of soft computing. MARS makes no assumptions about the original functional relationships that exist between a target variables and its related independent variables, thus this model is useful when there is an unknown relationship giving information on the parameters relevance from a high-dimensional panel data. Furthermore, in this study, MARS helps doing a selection of the variables for eliminating the presence of collinearity. In fact, MARS builds a model through the forward pass and the backward pass phases. The forward pass phase builds an overfit model characterized by a good fit to the data, and the backward pass phase eliminates the least effective terms finding the best sub-model [19]. The MARS ability to find relationships between variables can help explaining the relations with the target variable. In fact, the underground economy [4] contributes to hinder some relations linked to GDP in the analysis on the Italian case. MARS is also useful in limiting the multicollinearity problems thanks to its variable selection process. We must consider that in Italy there is the presence of a North-South dualism also regarding human capital quality [3], therefore this resource must be widely considered. For example, NEET rate is about 15–20% in the Center-North and 30–40% in some Southern regions, the workers involved in lifelong learning are respectively about 10% and 5% (Istat data). With the aim of comparing the wealthiest and the relatively poorest areas, *a priori* division of the regions in homogeneous group represents a fundamental step of the analysis [26]. However, we should find groups of spatially neighboring regions because the economic performance of a region affects the neighboring [14]. We apply MARS to the resulting groups and we compare the statistically significant determinants. What lacks in the less wealthy areas may be found among the reasons of the income divergence.

2 The Selection of the Income Determinants

The target variable considered in studies on regional inequality is GDP *per capita* (constant 2010 values, Istat data). Among the independent variables, we focus on the significant role played by the human capital. This type of capital represents one of the most important economic assets in the advanced economies [20] and it strongly influences the regional economic dynamics [16]. Furthermore, Abramo *et al.* [3] prove the presence of an Italian North-South dualism both in the research and in the educational system. Human capital is observed using educational [22] and lifelong learning data [13]. We consider the contribution of the principal economic sectors, through their added value contribution on the total, as Barro and Sala-I-Martin [5] consider, in a study of regional convergence, the contribution of each sector on GDP *per capita*. We also include data on the entrepreneurial vitality because the role of businesses is fundamental on the local development, despite not often used in studies on economic growth [27].

Of course, the international trade, observed with the export levels on GDP, represents one of the main supports to the aggregate demand, as Guerrieri and Iammarino [21] focus on the Italian regional specialization and diversification, that can represent a chance for the southern area through export. Finally, data on the financial sector must be consider. We select an Istat indicator representing the differential between the borrowing rates, calculated as the difference between the Centre-North and South, to represent the financial system dualism. This is due to the fact that the costs of borrowing increases during economic slowdown and in the worst performing areas (as Southern Italy) because, for example, of the increasing risk of default [11]. We present the list of variables (2004–2014) and the source of the data:

- population aged 25–64 with less than primary, primary and lower secondary education (M and F; levels 0–2 ISCED 2011, %) – Eurostat;
- population aged 25–64 with upper secondary and post-secondary non-tertiary education (M and F; levels 3–4 ISCED 2011, %) – Eurostat;
- population aged 25–64 with tertiary education (M and F; levels 5– ISCED 2011, %) – Eurostat;
- school dropout (M and F; % of population aged 18–24 with at most a sec. education, and who have not completed a training course (of more than 2 years) and who do not attend school courses or training) – Eurostat;
- NEET rate (M and F; young people neither in employment nor in education and training, %) – Eurostat;
- employed lifelong learning (employed 25–64 years engaged in training and education on 100 employed persons in the corresponding age group, %) – Istat;
- unemployed lifelong learning (M and F; unemployed 25–64 years engaged in training and education, %) – Istat;
- net enrollment rate in the Company Register (businesses registered minus the ceased ones on the total businesses registered in the previous year, %) – Istat;
- export/GDP ratio (%) – our elaborations on Istat data;
- differential in lending rates on loan facilities between the South and the Center-North of Italy; the lending rates are on total cash loans) – Istat;
- degree of use of PCs, Internet access, broadband availability and corporate website, for businesses with more than 10 employees – Istat;
- value added for agriculture, forestry and fishing (ratio of total value added) – our elaborations on Istat data;
- value added for quarrying, manufacturing, electricity supply, gas, steam and air conditioning, water supply, sewerage, waste management and remediation activities (ratio of total value added) – our elaborations on Istat data;
- value added for constructions (ratio of total value added) – our elaborations on Istat data;
- value added for wholesale and retail trade, repair of motor vehicles and motorcycles, transportation and storage, accommodation and food services, information and communication services (ratio of total value added) – our elaborations on Istat data;
- value added for financial and insurance activities, real estate, professional, scientific and technical activities, administrative and support services (ratio of total value added) – our elaborations on Istat data;

– value added for public administration and defense, social security, education, health and social work activities, arts, entertainment and recreation activities and other services (ratio of total value added) – our elaborations on Istat data.

Endogeneity problems can be present among some independent variables and the GDP *per capita*, however, our goal is not to explain the formation of the GDP but find the statistically significant variables.

3 Clustering of the Italian Regions

In order to examine the income determinants in the wealthiest and the poorest areas, the possible mobility of some regions in the studied period must be considered, staring from the official Istat grouping. We propose two clustering methods, K-means and Hierarchical. Considering the Eurostat NUTS2 level, we have data on the 19 regions and 2 autonomous provinces. The discriminating variables are 14 on human capital (the top 7 of the list for male and female) and GDP *per capita*, because we consider the inequality with a focus on education and lifelong learning.

In Fig. 1, we include Trentino-Alto Adige in the group of its most populous autonomous province, Trento. The results for the selected period, 2004–2014, show a dualism between the geographical North and South, as historically proven, and the two cluster techniques reveal two groups in the Center-North. Two central regions are in the South group in 3–b, but these regions have an average income very closer to the North. For this reason, we consider the K-Means division. The South group (white area in Fig. 1a) coincides with the official Istat *Mezzogiorno*, *i.e.* the Southern regions and major islands. This division confirms the historical dualism between the less wealthy area and the central and Northern regions (black and grey in Fig. 1a). We choose two representative subgroups of the richest area, not only for economic reasons (a difference is given by the fact that the black regions have the higher GDP *per capita*) but to have two possibilities of comparison. Furthermore, each group represents approximately one third of the Italian population considering these clusters (about 18, 22 and 21 million people).

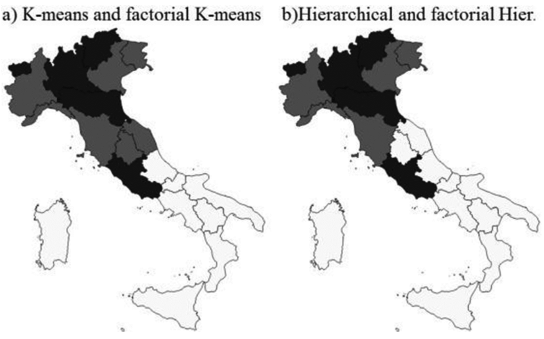

Fig. 1. Clustering of the Italian regions according to the human capital data and GDP per capita (2004–2014)

4 The MARS Model

MARS [18] is a non-parametric regression technique for solving regression-type problems. This model has no expectations on the underlying functional relationships between a target variable and the regressors. We apply two-sided truncated functions of the form $\pm(x-t)_+$ representing basis functions for linear or nonlinear expansions. Subsequently, with the aim of setting the coefficient values to fit the data in the most efficient way, the basis functions and the parameters (estimated via OLS) are combined to provide the predictions, assumed the inputs. The result is a geometrical procedure that is better than a standard approach. The multivariate splines algorithm derives from two-sided truncated functions of the predictors (x): $b_q^{\pm}(x-t) = [\pm(x-t)]_+^q$.

Basis functions are employed for generalizing spline fitting to higher dimensions. The multivariate spline basis function is (with two-sided truncated power basis for the univariate functions):

$$B_m^{(q)}(x) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+^q \tag{1}$$

with products involving the truncated power functions with polynomials of a lower order than q. These functions are the result of recursive partitioning and they are a subset of a complete tensor product ($q = 0$) spline basis with knots at every (distinct) marginal data point value [18]. The application of algorithms allows us to have the model:

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+ \tag{2}$$

in which a_0 is the coefficient of the constant basis function B_1, while the sum is over the basis functions B_m. A transformation of the model is useful to have a better predictive evidence on the relation between the response y and the covariates x.

4.1 The Results

We apply the MARS model to the three groups (see Fig. 1a). The comparison between the determinants useful in the wealthiest areas and the South must show deficiencies on which formulate policy. We present the MARS results considering no. 22 variables for the three groups (the first Center-North group is the black in Fig. 1a, while Center-North 2 is the grey group):

GDP_{CN1} = 29366,3498343309 − 3589,35951639091 * max (0; − Differential_lending_rates − 0,03970285894958) + 564,977344570989 * max (0; 17,330847237003 − NEET_F) + 647,622010522125 * max (0; VA_comm − 23,1695158977861) − 338,738840601931 * max (0; 23,1695158977861 − VA_comm) − 291,628095772613 * max (0; Edu_5–8_F − 14,9) + 317,2903450178 * max (0; 14,9 − Edu_5-8_F) − 95,6111400965899 * max

(0; 30,266629949516 − Export) + 286,096817645155 * max (0; Edu_3–4_M − 41,8) + 322,856140634558 * max (0; 41,8 − Edu_3-4_M) − 1310,59787295588 * max (0; Enrollment_buss − 1,15056059797117) − 609,475268577884 * max (0; Edu_3–4_M − 46,9) − 75,5817185069516 * max (0; Broadband − 70,3460464008)

GDP_{CN2} = 32552,5479458084 − 181,177256150065 * max (0; 16,9 − Edu_5–8_F) + 705,807181891711 * max (0; VA_agr − 2,46543766459935) + 3495,79573171618 * max (0; 2,46543766459935 − VA_agr) − 199,697609189166 * max (0; VA_pubbl − 22,08488708359) − 300,423449225602 * max (0; NEET_M − 12,979390509693) − 497,853182545221 * max (0; VA_finance − 26,8758306675439) + 768,032672652473 * max (0; 26,8758306675439 − VA_finance) − 177,587312671335 * max (0; Dropout_F − 15,335292546086) + 449,159846179549 * max (0; Empl_lifelong_M − 6,81555362821625) − 24,309612064087 * max (0; 80,8871595330739 − Broadband) − 218,746529253119 * max (0; 46,1 − Edu_3–4_F) − 504,163985425754 * max (0; Unempl_lifelong_F − 6,41037014932596) + 383,086153331738 * max (0; 6,41037014932596 − Unempl_lifelong_F)

GDP_S = 17071,1322439226 + 131,320774988255 * max (0; 33,5309920231828 − NEET_F) + 155,796598835784 * max (0; Dropout_F − 18,9348238394318) − 342,820844783344 * max (0; Edu_0–2_F − 53,1) + 216,444004595303 * max (0; 53,1 − Edu_0–2_F) − 74,4855093336676 * max (0; NEET_M − 19,1838437206716) + 273,30143729301 * max (0; 19,1838437206716 − NEET_M) + 263,083417772284 * max (0; Edu_0–2_M − 46,8) − 151,55129092139 * max (0; VA_comm − 19,677182384134) + 443,382569327129 * max (0; 19,677182384134 − VA_comm) − 338,422499845198 * max (0; VA_finance − 23,3979742801912) − 91,286233129198 * max (0; 23,3979742801912 − VA_finance) + 250,627730663668 * max (0; Empl_lifelong_F − 7,73188152439938) + 708,175041261683 * max (0; 3,99276970915113 − VA_agr)

The results show differences between the two groups of the Center-North. The first one is based on the strength of manufacturing businesses in exporting abroad and obviously the group needs the credit leverage and a trained human capital. The second one encloses the Northern regions relatively more affected by the crisis, however, the added value of different sectors is important, but also the continuous training to qualify workers and retrain the unemployed. Even in this case the human capital plays an important role. In the South, the "useful" human capital is not present, and NEET and school dropout are social problems. It is significant the presence, albeit in subsequent knot of the model, of different sectors whose added value is statistically significant. For having a focus on education and vocational training, we also present MARS result on the solely human capital variables for the three groups:

GDP_{CN1} = 28040,8757383946 + 281,958908864165 * max (0; 19,2 − Edu_5–8_F) + 692,033056856368 * max (0; 17,330847237003 − NEET_F) − 1531,92671671789 * max (0; Edu_3–4_M − 46,9) − 1312,87441845612 * max (0; 4,90952417557076 − Empl_lifelong_M) + 664,422864118961 * max (0; Edu_3–4_F − 42,2) − 311,470529768982 * max (0; NEET_M − 12,392450034517) − 355,391872361449 * max (0; Unempl_lifelong_M − 7,82242927248468)

GDP_{CN2} = 55961,7225821171 − 7117,07669234711 * max (0; Edu_5–8_F − 16,9) + 6428,40611739239 * max (0; 16,9 − Edu_5–8_F) + 571,171256171591 * max (0; Unempl_lifelong_M − 9,95934959349594) − 262,481927561333 * max (0; 9,95934959349594 − Unempl_lifelong_M) + 1459,70787195025 * max (0; 5,59284116331096 − Unempl_lifelong_F) − 6472,49939488940 * max (0; Edu_0–2_F − 37,2) + 5975,92944332297 * max (0; 37,2 − Edu_0–2_F) + 696,174280430151 * max (0; Empl_lifelong_M − 6,81555362821625) + 722,070000598687 * max (0; Edu_5–8_M − 15) − 6383,39245658321 * max (0; Edu_3–4_F − 42,6) + 6607,32344887727 * max (0; 42,6 − Edu_3–4_F) + 641,953599419053 * max (0; 14,294401061366 − NEET_F)

GDP_{S} = 17085,4008884516 + 220,661529522481 * max (0; 33,5309920231828 − NEET_F) + 327,869417021748 * max (0; Dropout_F − 18,9348238394318) − 90,3335055675469 * max (0; NEET_M − 25,8345420570518) + 228,279499242398 * max (0; 25,8345420570518 − NEET_M) + 353,520059663269 * max (0; Edu_3–4_F − 38,6) − 321,184170529905 * max (0; Edu_0–2_F − 52,6) + 119,724901898192 * max (0; Edu_0–2_M − 46,8) − 575,788189577109 * max (0; 3,89405339219256 − Empl_lifelong_M) − 314,531903433760 * max (0; 7,70803092691886 − Unempl_lifelong_M)

This detailed focus confirms that, in the Southern area, early school dropout and unemployment afflict the local society damaging one of the most important long-term resources. In the last knots, values on vocational training (men only) are relevant, while completely lacks the relevance of secondary and tertiary education, which instead are confirmed (and in some cases in the first knots) in the other two groups (and for both genders, thus confirming more equality on education and employment opportunities).

5 Policy Suggestions and Conclusions

In this study, we investigate and compare the different income determinants in three cluster of Italian regions. The aim is to observe the deficiencies of the less wealthy area and most affected by the 2007 crisis, identified with the so-called *Mezzogiorno*. Our focus is on the level of human capital, seen as education and vocational training. We apply a MARS model to the three groups to search for the statistically significant variables. The results show that the strengths of the North refer to businesses capacity in export thanks to several local characteristics, as educated workers and a more efficient

financial system. This is missing in the South, plagued by school dropout, unemployment and young NEET problems (see Sect. 1). Of, course, also in the North we would expect a greater contribution of human capital, while in the south only primary education is relevant, as in Di Liberto [12]. The poor economic vitality in the South discourages educated workers, forcing them to migrate toward the northern regions or abroad [7], enhancing regional inequality being a skill-selective migration [17]. The mobility of workers between the southern regions could be a short-term solution but there are deficiencies in the exchange of labor information required in this case [15]. Obviously, in the long term, incentives and facilities should be provided for businesses, which in the past had shown competitiveness with local specializations since the European integration [6]. The education system should also be improved, up to the universities [3], but for this purpose we should first change the social and cultural substrate, towards a model that encourages and rewards investment in advanced education.

References

1. A'Hearn, B., Venables, A.J.: Regional disparities: internal geography and external trade. In: Toniolo, G. (ed.) The Oxford Handbook of the Italian Economy Since Unification. Oxford University Press, Oxford (2013)
2. Abraham, A., Steinberg, D., Philip, N.S.: Rainfall forecasting using soft computing models and multivariate adaptive regression splines. IEEE SMC Trans. Spec. Issue Fusion Soft Comput. Hard Comput. Ind. Appl. **1**, 1–6 (2001)
3. Abramo, G., D'Angelo, C.A., Rosati, F.: The North–South divide in the Italian higher education system. Scientometrics **109**(3), 2093–2117 (2016)
4. Ardizzi, G., Petraglia, C., Piacenza, M., Turati, G.: Measuring the underground economy with the currency demand approach: a reinterpretation of the methodology, with an application to Italy. Rev. Income Wealth **60**(4), 747–772 (2014)
5. Barro, R.J., Sala-I-Martin, X.: Convergence across states and regions. Brookings Papers Econ. Act. **22**(1), 107–182 (1991)
6. Basile, R., Giunta, A., Nugent, B.: Internationalisation process of small and medium sized firms in Italy over the nineties. In: Pietrobelli, C., Sverrisson, A. (eds.) Linking Local and Global Economies. Routledge, London (2004)
7. Biagi, B., Faggian, A., McCann, P.: Long and short distance migration in Italy: the role of economic, social and environmental characteristics. Spat. Econ. Anal. **6**(1), 111–131 (2011)
8. Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., Pancotto, F.: Amoral familism, social capital, or trust? the behavioural foundations of the Italian North-South divide. Econ. J. **126**(594), 1318–1341 (2016)
9. Daniele, V., Malanima, P.: Il prodotto delle regioni e il divario Nord-Sud in Italia (1861–2004). Rivista di Politica Economica **97**(2), 267–316 (2007)
10. Daniele, V., Malanima, P.: Il divario Nord-Sud in Italia 1861–2011. Rubbettino, Soveria Mannelli (2011)
11. Del Giovane, P., Nobili, A., Signoretti, F.M.: Supply tightening or lack of demand? An analysis of credit developments during the Lehman Brothers and the sovereign debt crises. Banca d'Italia - Temi di discussione, 942 (2013)
12. Di Liberto, A.: Education and Italian regional development. Econ. Educ. Rev. **27**(1), 94–107 (2008)

13. Edwards, R., Ranson, S., Strain, M.: Reflexivity: towards a theory of lifelong learning. Int. J. Lifelong Educ. **21**(6), 525–536 (2002)
14. Ertur, C., Le Gallo, J., Baumont, C.: The European regional convergence process, 1980–1995: do spatial regimes and spatial dependence matter? Int. Reg. Sci. Rev. **29**(1), 3–34 (2006)
15. Faini, R., Galli, G., Gennari, P., Rossi, F.: An empirical puzzle: falling migration and growing unemployment differentials among Italian regions. Eur. Econ. Rev. **41**(3–5), 571–579 (1997)
16. Fleisher, B.M., Li, H., Zhao, M.Q.: Human capital, economic growth, and regional inequality in China. J. Dev. Econ. **92**(2), 215–231 (2010)
17. Fratesi, U., Percoco, M.: Selective migration, regional growth and convergence: evidence from Italy. Reg. Stud. **48**(10), 1650–1668 (2014)
18. Friedman, J.H.: Multivariate adaptive regression splines. Ann. Stat. **19**(1), 1–67 (1991)
19. Friedman, J.H.: Fast MARS. Department of Statistics. Stanford University Technical Report, Stanford (1993)
20. Goldin, C.: Human capital. In: Diebolt, C., Haupert, M. (eds.) Handbook of Cliometrics. Springer, Heidelberg (2016)
21. Guerrieri, P., Iammarino, S.: Dynamics of export specialization in the regions of the Italian Mezzogiorno: persistence and change. Reg. Stud. **41**(7), 933–948 (2007)
22. Hanushek, E.A.: Economic growth in developing countries: the role of human capital. Econ. Educ. Rev. **37**, 204–212 (2013)
23. Iuzzolino, G., Pellegrini, G., Viesti, G.: Regional Convergence. In: Toniolo, G. (ed.) The Oxford Handbook of the Italian Economy Since Unification. Oxford University Press, Oxford (2013)
24. Piffer, D., Lynn, R.: New evidence for differences in fluid intelligence between North and South Italy and against school resources as an explanation for the North–South IQ differential. Intelligence **46**(5), 246–249 (2014)
25. Sephton, P., Mann, J.: Further evidence of an environmental kuznets curve in Spain. Energy Econ. **36**(C), 177–181 (2013)
26. Terrasi, M.: Convergence and divergence across Italian regions. Ann. Reg. Sci. **33**(4), 491–510 (1999)
27. Van Stel, A., Carree, M., Thurik, R.: The effect of entrepreneurial activity on national economic growth. Small Bus. Econ. **24**(3), 311–321 (2005)