How to Best Classify Rural and Urban?

Edoardo Pizzoli
Food and Agriculture Organization of the United Nations (FAO)
Rome, Italy
edoardo.pizzoli@fao.org

Xiaoning Gong
Food and Agriculture Organization of the United Nations (FAO)
Rome, Italy
xiaoning.gong@fao.org

Abstract: While "agricultural and rural development" is one of the key policy areas, there is no universally accepted way on how to classify the urban and rural. On one hand, dissatisfied with the OECD approach in which a "rural area" is defined merely based on the population density, there has been a rich literature in the statistical discourse and policy debate on what other variables should be considered and included for the urban-rural typology. On the other hand, according to the rural areas observed in OECD countries, where the population has migrated to the urban and agricultural production has become less important, some people argue that the traditional way of thinking "rural" as the same of "agricultural" is not true anymore for the developed countries and the same transformation is going on in the developing countries. The salient feature of this paper is to apply an econometric logistic regression model to two datasets for Italy, an OECD country and China, a non-OECD country. The results of the model unambiguously demonstrate that, first, to classify the rural and urban, a multidimensional approach by taking into account both economic activities and geographic dimension along with population density is more appropriate; and second, agriculture is still highly relevant to the rural, and thus an important determinant factor in defining the rural area.

KEYWORDS: Agriculture and Rural Development, Classification, Logit Model, Rural and Urban Typology, Population Density.

1. Introduction

Agricultural and rural development is one of the key policy areas. As a specialized UN agency, FAO has focused special attention on developing rural areas, where most of world's poor and hungry people are living, in order to carry its mandate to "raise the levels of nutrition, to improve agricultural productivity and to increase the living conditions of rural populations." In many FAO documents rural areas and rural development are associated with areas where there is agricultural activity and a relevant percentage of total population is employed in the sector. The terms of rural and agricultural are considered interchangeable.

Two main trends observed in OECD countries during the last decades are an overall enlargement of urbanization resulting from population migration from rural to urban, and a shift of rural economies from agriculture to manufacturing, and then from manufacturing to services as the leading sectors of activity (OECD, 1994). Applying OECD rural-urban definition, one would find more than half of the production in rural area is generated by the services sectors while agriculture contributes to much less (see Table 5 in the Appendix). At the same time, globalization has also affected the fundamental structure of the economy and geographical specialization, including the relationship between the rural and urban. Consequently, in OECD countries there has been an effort to find out a better and appropriate territorial classification in term of urban and rural areas (Sotte, 2003; UNECE, 2005).

There is a doubt that the traditional way of thinking "rural" as the same of "agricultural" may not be true anymore for developed countries (Bollman, 2007) and the same transformation is going on in several developing ones (UNECE, 2005); despite the fact that while in the large developing countries, such as Brazil, China, and India, however, agriculture still plays an important role in their whole economy in terms of employment and income. The similar situation can be found in many other countries in all the continents (Table 6). Thus, for many non-OECD countries, rural areas are still associated with agricultural activity and coincide with agricultural areas.

In this paper, we will apply an econometric logistic regression model to explore if agriculture and other variables are still relevant in the rural urban typology and to propose a multidimensional approach to the classification of urban and rural areas.

The rest of this paper is organized as follows. The next session is a literature review of "rural" definitions in rural and urban classifications. Through a critical review of the OECD definition developed in the 90's and other literature to provide guidance on what variables to be considered for the following empirical study of Italy and China. It is followed by a quantitative analysis. First is to verify if the population density criterion proposed by OECD is sufficient to identify the rural areas; and second, to test how significantly other variables suggested by the literature contribute to the urban-rural typology. Data collected for two countries, Italy and China are used as a case study. The last section summarizes and concludes the paper.

2. Literature Review

At the early 90's OECD launched a work on sub-national territorial statistics and indicators for member countries as a part of a Rural Development Programme. In its first report, *Creating Rural Indicators for Shaping Territorial Policies* (OECD, 1994), a basic conceptual framework is provided to subdivide the territory of member countries in three hierarchical levels of geographical detail: national, regional and local. Within each of its 30 member countries, local areas (70.000 units) were clustered according to the rural and urban typology, where rural areas was identified as communities with population densities below 150 inhabitants per square kilometer. The regions are also grouped in three clusters depending on the share of regional population living in rural areas: predominantly rural (over 50 %), intermediate (15 to 50 %) and predominantly urbanized (below 15 %).

Some countries, such as Japan, would require different population threshold to identify rural areas, for example, a threshold of 500 inhabitants per square kilometer (km²), due to the higher population density compared with others. As well as Belgium, where two different optimal population density limits are identified for its two regions: at 600 inhabitants per km² for Flanders and at 300 inhabitants per km² for Wallonia (Lenders, 2007).

Other countries consider that even municipalities are too large geographical areas and a thinner territorial subdivision is necessary to identify rural areas, such as the EUROSTAT work on geo-referencing of statistical data, where a modified OECD definition is applied on EU grid cells (EC, 1999; Vard, 2005; Albrecht, 2006). The third group would like to extend the classification from just a binomial rural / urban typology to a more various and complex territorial clustering. An example can be found in European Union territorial classification where three degrees of urbanization are defined for areas at NUTS 3 levels: densely, intermediate and sparsely populated (EUROSTAT, 2007).

Another string of development is to introduce more than one criterion or other variables for territorial classification, in addition or in alternative to the population density. An example is the UK national statistics where 8 rural / urban types, of which 6 defined as rural, are applied on census output areas (DEFRA, 2005; Bibby, 2005).

<u>Economic activities including agriculture</u>: Taking into account of the degree of economic development: high degree of specialization in agriculture, in “traditional” rural areas, or a high dispersion of economic activities with a relevant weight of agriculture and services sectors, in “modern” rural areas. To describe the economic profile of the area, employment, production, value added, and land use by sectors are considered (Cecchi, 1999; Bryden, 2001; FAO, 1986, 1993, 2005).

<u>Socio-economic structural characteristics</u>: Some other studies suggest the use of socio-economic structural characteristics such as, infrastructures (Plessis, 2001; Albrecht, 2006) or human capital (Cecchi, 1999). The educational level of the population in the area is one of the measurements to identify the human resources and skills. In rural areas they are expected to be low so that mainly manual activities are developed.

<u>Spatial dimension of social organization</u>: indicators refer to the spatial dimension of social life, highlighting that in rural areas distances are higher and opportunities are lower with respect to urban areas (Plessis, 2001; UNECE, 2005; Albrecht, 2006). That will result in fewer services available. In this case to find out a suitable and available variable is not always easy, which is why several alternative solutions are applied by the authors suggesting this approach. Distances from some key services, services available per square kilometer or per capita, and the length of roads for square kilometers are some of the proxy variables.

<u>Natural characteristics</u>: Some of the studies propose to consider the natural characteristics of the earth surface and the environment. An example is a European Joint Research Centre

(JRC) study on territorial classification that considers land cover profiles (arable, forest, etc.), topographic roughness (mountain, hill, plain) and climate relevant (Gallego, 2004; Vard, 2005). Rural areas should be mainly covered by agricultural land, forests and natural areas, including disadvantaged areas for human activity (mountain, extreme climate conditions, etc.). The idea behind is that natural environment significantly differentiates human opportunities and its behavior and, with this respect, rural and urban areas should result deeply different.

3. Empirical Study

In this section, we will apply those variables suggested by the literature as reviewed above to an empirical study in two countries: Italy and China. Using these variables as explanatory ones, and the rural / urban characterization of areas is considered as an endogenous qualitative variable in a logistic regression model to test if they are significant or not in explaining rural and urban differences in considered administrative areas and thus they can be used as criteria for categorize the urban and rural. If these variables are statistical significant, it implies that a rural urban typology made solely based on the population density as suggested by the OECD definition is not properly done, as it has omitted and not taken into account other important affecting factors.

3.1. Model

The model specified in this paper to study the relationship between the rural / urban classification and several suggested indicators is a probability model with a discrete dependent variable. This choice is appropriate because the classification, assumed as dependent in the model, is a binary variable with two qualitative outcomes: rural and urban; while the predictors are the factors discussed in the preceding section (Maddala, 1983).
In this framework, the probability of an area to be rural is linked to the set of factors.
A suitable assumption for the right-hand side of the model is the logistic distribution:

(1) **Logit model :** $Prob[Y=1] = \Lambda(\beta' x),$
$Prob[Y=0] = 1 - \Lambda(\beta' x).$

Where:

***Y** is a binomial variable (**Y=1** if rural, **Y=0** if urban);*
***x** is the vector of explanatory variables;*
***β** is the vector of parameters for **x**;*
***Λ(.)** is the logistic cumulative distribution function.*

The vector of ***β*** reflects the impact of changes in ***x*** on the probability of ***Y***, that means the impact of explanatory indicators on the probability of an area to be rural or urban.

The ***β*** parameters are estimated on Italian and Chinese data and the ***x*** explanatory variables are chosen among the suggested proxy variables subject to availability.

In two steps, the first is to estimate a model with just population density and then other significant variables are introduced.

The logistic regression model is estimated using maximum likelihood algorithm. An efficient way to assess the goodness-of fit of logistic regression is to use pseudo $\boldsymbol{R^2}$. As the specified model is not linear, to understand the marginal effect of changes of $\boldsymbol{x}$ on rural probability ***Prob[Y=1],*** the partial derivatives will be computed at the means of the regressors.

3.2. Data Sets

Data employed in the study are from official national sources, the National Institute of Statistics (ISTAT) for Italy and the National Bureau of Statistics (NBS) for China. A first determining choice is on the territorial subdivision to use that mainly depends on national administrative levels and detail of data available. A special focus is necessary to clarify the relationship between the administrative territorial subdivision and rural areas, as geographical entities.

3.2.1. Number of Observations

Italy

In Italy are available three hierarchical levels of administrative subdivision: 1) Regions; 2) Provinces; 3) Municipalities that is the smallest in terms of territorial extension. More than one geographical context (cities, agricultural lands, mountains, valleys, costs, etc.) can be found in any municipality.

The Italian administrative subdivision is harmonized with the 3 levels of EU classification: Nomenclature of Territorial Units for Statistics (NUTS) 1, 2 and 3 (EUROSTAT, 2005). NUTS 3 corresponds to Provinces level for Italy; that is why it is considered less suitable than the national municipalities level for rural and urban classification.

In this study the 8115 municipalities are classified 2551 as rural, on the base of the European Union indications for the Structural Funds policy (EC, 2005) and the remaining as urban.

China

The Constitution of the People’s Republic of China stipulates that the administrative areas in China are divided as: 1) The whole country is divided into provinces, autonomous regions and municipalities directly under the central government; 2) Provinces and autonomous regions are divided into autonomous prefectures, counties, autonomous counties and cities; 3) Autonomous prefectures are divided into counties, autonomous counties and cities; 4) Counties and autonomous counties are divided into townships, nationality townships and towns; 5) Municipalities and large cities are divided into districts and counties.

The 1576 counties and 374 cities can be the reasonable level to study the rural and urban classification. Counties could be labeled as rural, while cities by definition as urban.

3.2.2. Variables

- Population density is available from both datasets;

- Agricultural land or agricultural employment is selected to reflect the importance of the agricultural sector. The ratio of the agricultural employment in the total employment is used as a proxy of economic specialization;

- Population at fixed level of education on total population is used to measure human resources and skills;

- Telephone per capita is chosen, as services available, to approximate that in rural areas distances are higher and opportunities are lower with respect to urban areas;

- The ratio between "green" land (agricultural, forest and natural areas included) of the total land is used to refer to the physical geographical characteristics of the area: soil roughness, land cover and climate.

3.2.3. Statistical Analysis

Two statistical properties are tested for selected variables: variability and correlation. Variability over areas is a necessary condition for a variable to be an explanatory one in the model, while a high correlation coefficient between two variables indicates that one could be excluded.

Italy

Table 1 – Standard Deviation over Mean and Correlation Matrix of Proxy Variables for Italian Municipalities (*2001*)

Variable	Std. Dev. / Mean	Correlation Coefficients					
		pop_d	*s_agr*	*emp_agr_d*	*s_emp_agr*	*s_edu*	*s_green*
pop_d	2.20	1	-0.08	0.55	-0.25	0.21	-0.19
s_agr	0.6		1	0.19	0.27	-0.08	0.76
emp_agr_d	1.4			1	0.16	0.01	0.04
s_emp_agr	0.9				1	-0.46	0.25
s_edu	0.2					1	-0.09
s_green	0.5						1

Source: Istat

Note:
pop_d population for km^2 (number);
s_agr share of agricultural used land over total land (%);
emp_agr_d agricultural employment for km^2 (number);
s_emp_agr share of agricultural employment over total employment (%);
s_edu share of educated population (secondary school) over total population (%);
s_green share of "green" land (agriculture, forests and natural areas) over total land (%).

The table above shows that the population density has the highest variability over Italian municipalities, and thus it can be a good candidate for territorial classification. The variability of other variables is relatively lower.

In general, the correlation coefficients are not high among the selected variables. Especially, the population density in Italy is disjointed with the agricultural used land that rooted in agriculture. The population density is much correlated with the share of agricultural employment on total employment in the area. A very strong positive relationship is found between the shares of agricultural used land and share of green land over total land. This is a clear indication that most of green land in Italy is employed in agriculture. At last, a significant negative correlation is found between the share of agricultural employment and education level. Greater is specialization in agriculture in the area, lower is education level.

China

Table 2 – Standard Deviation over Mean and Correlation Matrix of Proxy Variables for Chinese Counties and Cities (*2005*)

Variable	Std. Dev. / Mean	Correlation Coefficients		
		pop_d	*s_emp_agr*	*s_tel*
pop_d	0.95	1	0.08	0.17
s_emp_agr	0.66		1	-0.50
s_tel	0.72			1

Source: NBS

Note: s_tel share of population with a telephone line (%).

Over Chinese counties and cities the highest variability is found for population density, as in the Italian case, but the absolute value is much lower. This result could be explained on the difference in territorial extension of the two administrative levels: more extended areas in China will include similar population levels. Furthermore, the variability of the other two variables, share of agricultural employment and share of telephone lines on the population, is not much different. All of them are good candidate as classification criteria.

In terms of correlations, the population density has a very low relationship with the other two variables and so it has a limited capability to represent them as a proxy variable. Low populated areas in China could be or not specialized in agriculture and with a high or low availability of communications infrastructure. It is important to notice that it is possible to find high populated areas specialized in agriculture. The correlation coefficient is even slightly positive, at the opposite of the Italian case. At the end, the share of agricultural employment (agriculture specialization) and the share of telephone lines over population (services availability) are negatively correlated. Both of them should be able to grasp further structural variability of the areas.

3.3. Results

The results from the estimation of four Logit models are shown in the following table:

Table 3 – Comparison of the Estimated Logistic Models: Data for Italy and China

Variable	Italy		China	
	Model (1.1): only population density	Model (1.2): more variables	Model (2.1): only population density	Model (2.2): more variables
pop_d	-0.027 (0.00)	-0.042 (0.00)	-0.002 (0.00)	-0.002 (0.00)
s_agr		3.438 (0.00)		
s_emp_agr		18.665 (0.00)		0.238 (0.00)
s_edu		-4.156 (0.00)		
s_green		-1.015 (0.00)		
s_tel				-5.587 (0.00)
Constant	1.711 (0.00)	0.555 (0.02)	2.050 (0.00)	1.819 (0.00)
Pseudo R^2	0.393	0.612	0.045	0.235

Note: P>z in parenthesis.

The marginal effects of changes in the explanatory variables are calculated as partial derivatives of models (1.2) and (2.2).

Table 4 - Marginal Effect of a Change in Regressors on the Probability of an Area to Be Rural, calculated at the Mean of the Regressors

Variable	Marginal Effect	
	Italy	China
pop_d	-0.00003	-0.00024
s_agr	0.00271	
s_emp_agr	0.01472	0.02803
s_edu	-0.00328	
s_green	-0.00080	
s_tel		-0.65699

The results show:

- All the selected variables are statistically significant which are consistent with the literature;
- The population density alone is significant but its marginal effect is quite small and the goodness of fit is poor.
- The introduction of further variables strongly increases the overall fitting of the model: the Pseudo R^2 is much higher in models (1.2) and (2.2) with respect to (1.1) and (2.1). They contribute to a better specification of the model.

- Agriculture matters in both countries, in terms of economic specialization (share of employment) and also, in the Italian case, in term of historical roots (share of agricultural land). It is interesting to find that in both countries this characteristic seems to be much more relevant than population density.

- Education level of the population and availability of services (telephone lines) have a negative effect on the probability of an area to be rural, greater than population density.

4. Summary and Conclusion

The econometric model proposed in this paper turn out to be very fruitful, as it is the first attempt to study this classification issue on an empirical ground, comparing the different proxy variables suggested by the literature. The results from the empirical study of Italy and China show that the population density itself is not a sufficient criterion to classify the rural and urban. There are other variables, such as agriculture and economic specialization, human resources and skills, land cover and spatial dimension of social life, that are also relevant. A joint use of them would significantly improve the accuracy in the approximation of areas' probability to be rural or urban. This result suggests that we should consider applying a "multidimensional" approach for the typology of urban and rural. Agriculture, in the sense of historical roots or economic specialization, plays an important role in rural areas identification. This is for both country cases: Italy and China. While for China, a developing country where "traditional" rurality, which is with areas specialized in agriculture, is still the normal case and population density is very often high; it seems to be surprised this result applies to Italy too; a developed OECD country where "modern" rurality, which is with areas with differentiated activities, is well developed in main regions of the country.

References

Albrecht, W. (2006) *Experiences with Grids at Eurostat*, Nordic Forum for Geo-statistics, Kongsvinger, Norway.

Bibby, P., Shepherd, J. (2005) *Developing a New Classification of Urban and Rural Areas for Policy Purposes – the Methodology*, DEFRA, London, UK.

Bollman, R. (2007) *The Demographic Overlap of Agriculture and Rural*, Agriculture and Rural Working Paper, No. 81, Statistics Canada, Ottawa, Canada.

Bryden, J. (2001) *Section 3: Rural Development,* in Proposal on Agri-Environmental Indicators (PAIS), Landsis, Luxembourg.

Cecchi, C. (1999) *Contextual Knowledge and Economic Exclusion in Rural Local Systems*, How to be Rural in Late Modernity, European Society for Rural Sociology XVIII Congress, Lund, Sweden.

DEFRA (2005) *Rural and Urban Area Classification 2004. An Introductory Guide*, London, UK.

EC (1999) *CORINE Land Cover. Technical Guide*, GI&GIS, Luxemburg.

EC (2005) *On Support for Rural Development by the European Agricultural Fund for Rural Development (EAFRD)*, Council Regulation, No. 1698, Brussels, Belgium.

EUROSTAT (2005) *Nomenclature of Territorial Units for Statistics – NUTS. Statistical Regions of Europe*, Website, Luxemburg.

EUROSTAT (2007) *European Regional and Urban Statistics. Reference Guide*, Luxemburg.

Fanfani, R., Brasili, C. (2003) *Regional Differences in Chinese Agriculture: Results from the 1997 First National Agricultural Census*, The Journal of Peasant Studies, Vol. 30, No. 3-4, Routledge, London, UK.

FAO (1986) *Population and the Labor Force in Rural Economies*, Economic and Social Development Paper, No. 59, Rome, Italy.

FAO (1993) *Rural Poverty Alleviation. Policies and Trends*, Economic and Social Development Paper, No. 113, Rome, Italy.

FAO (2005) *World Programme for the Census of Agriculture 2010.* A System of Integrated Agricultural Censuses and Surveys, Rome, Italy.

FAO (2007) *What is FAO?,* Website, Rome, Italy.

Gallego, F. J. (2004) *Mapping Rural/Urban Areas from Population Density Grids*, Institute for Environment and Sustainability, JRC - EC, ISPRA, Italy.

ISTAT (2006) *Atlante Statistico dei Comuni*, Informazioni, N. 25, Rome, Italy.

Lenders, S., Lauwers, L. and Kerselaers, E. (2007) *Rural-Urban Delimitation. A Statistical Analysis*, ILVA, Merelbeke, Belgium.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, Cambridge University Press, Cambridge, UK.

MIPAF (2006) *Piano Strategico Nazionale per lo Sviluppo Rurale*, Roma, Italy.

National Bureau of Statistics (2005) *2005 China County Statistical Yearbook*, China Statistics Press, Beijing, China.

OECD (1994) *Creating Rural Indicators for Shaping Territorial Policies*, OECD Publications, Paris, France.

Plessis, V., Beshiri, R. and Bollman, R. (2001) *Definitions of "Rural"*, Statistics Canada, Ottawa, Canada.

Sotte, F. (2003) *An Evolutionary Approach to Rural Development. Some Lessons for the Policymaker*, Associazione Alessandro Bartola, Collana Appunti, No. 3, Ancona, Italy.

Sotte, F. (2004) *From CAP to CARPE: the State of the Question*, 87° EAAE Seminar, Vienna, Austria.

Vard, T., Willems, E., Lemmens, T. and Peters, R. (2005) *Use of the CORINE Land Cover to Identify the Rural Character of Communes and Regions at EU Level,* in Trends of some Agri-environmental Indicators in the EU, Report of the European Commission, Brussels, Belgium.

UNECE, FAO, OECD, World Bank (2005) *Rural Household's Livelihood and Well-Being: Statistics on Rural Development and Agriculture Household Income*, Handbook, Geneva, Switzerland.

UN (2004) *World Urbanization Prospects: The 2003 Revision*, New York, USA.

Appendix

Table 5 - Urban on Rural Population and Share of Agricultural GDP Trends in OECD Countries (%)

OECD Countries	Years		
	1979-1980	1989-1991	1999-2001
	Urban on Rural Population		
US	2.8	3.0	3.8
non US	2.0	2.4	2.7
	Share of Agricultural GDP		
US	1.0	1.0	1.0
non US	4.9	4.3	3.8

Source: UN, FAO

Table 6 – Rural Population and Weight of Agriculture in Non OECD Countries (*1999-2001*)

Area	Estimated Rural Population	Agricultural Population		Share of Agricultural GDP in the	
				World	National Economy
	cumulative %	*cumulative %*	*% on rural pop.*	*cumulative %*	*% on total GDP*
Asia					
China	25.5	33.2	106.6	15.9	13
India	48.4	54.4	71.4	25.3	23
Indonesia	52.2	58.0	78.0	27.6	16
Bangladesh	55.5	61.0	69.1	28.6	25
Pakistan	58.5	63.8	71.7	30.2	24
Africa					
Nigeria	60.5	65.3	57.8	31.3	28
America					
Brazil	61.5	66.4	93.7	34.4	6
Non OECD Countries	90.8	96.6	84.0	56.0	26
World	100.0	100.0	79.0	100.0	4

Source: UN, FAO

Table 7 - Proxy Variables in Rural Classifications

Characteristic	Modality	Dataset	Index	Variable
1. Population with respect to territory	Lower with respect to urban areas, positive with respect to unmanned	Population	Population density (under threshold)	Inhabitants per square kilometers (%)
			Total population (under threshold)	Total population (number)
			Rural population (not urban)	Population outside urban centers (number)
2. Agricultural sector	Historically rooted on agriculture	Agricultural lands	Land cover profile (predominantly arable)	Arable land on total territory (%)
				Agricultural total area on urban area (>=< 1) Agricultural land on total territory (%)
		Employment in agriculture	Employment in agriculture (over threshold)	Employment in agriculture over areas (number)
3. Economic specialization	Higher degree of specialization in agriculture or greater dispersion over activities with respect to urban areas	Employment by sectors of activity	Share of agricultural employment on total employment (over a threshold)	Ratio of agricultural employment on total employment (%)
				Localization index (>=<1)
			Distribution of employment into economic activities (dispersion over activities)	Variance (> 0)
				Specialization index (0 - 1)
4. Human resources and skills	Lower skilled and educated labor force with respect to urban areas and specialization on manual works	Population by areas	Education level of population (low, under urban level)	Educated population on total population (%)
5. Spatial dimension of social organization	Higher distances and lower availabilities than urban areas	Services and communications (beds in hospital, telephone lines, roads, ...)	Accessibility (difficult, higher than urban areas)	Telephone lines per capita (number), others
6. Area's surface	Green and used by agriculture	Altimetry	Altimetry (high)	Average altimetry class (mountain, hill and plane)
		Land cover profile	Agricultural land, forests and natural areas (Over threshold)	Agricultural land, forests and natural area on territory (%)