REVENUES AND EXPENDITURES IN ITALIAN MUNICIPALITIES: A MULTILEVEL LATENT CLASS APPROACH[1]

Roberta Varriale, Domenico Passante, Cecilia Vitiello

Abstract. The statistical register for Public Administrations (REPA) contains information on structural and economic variables on a subset of the Italian PA Institutions. REPA includes different subpopulations, we focus on the subpopulation of municipalities, that are naturally grouped in regions. The economic variables of REPA include accrual and cash values, for both revenues and expenditures. This paper aims to analyse the economic management behaviour of Italian municipalities. To this aim, we adopt a multilevel latent class approach, which allows to simultaneously cluster municipalities and regions in groups with some typical profile starting from the observed economic variables.

1. Introduction

The statistical Register of Public Administrations (REPA) is a new product of the Italian national institute of statistics (Istat). REPA is integrated into a single logical environment, the Italian Integrated System of Statistical Registers (ISSR) (Luzi *et al.*, 2019), comprising a series of statistical registers (basic, thematic and extended) that centralise and integrate data from administrative sources, statistical surveys carried out by the institute and new and emerging sources for the ongoing production of official statistics. REPA contains information on structural and economic variables for the subset of Italian Public Administrations in the so-called "S13 list" produced by Istat[2], classified in different subpopulations.

In this paper, we consider the subpopulation of Italian municipalities included in the S13 list, naturally grouped in regions, and analyse their economic management behaviour, both in terms of revenue and expenditure. We use a multilevel latent class analysis approach introduced by Vermunt (2003), which allows us to simultaneously classify municipalities and regions into groups with a typical profile based on the information provided by some observed continuous indicators. The structure of the paper is as follows. Section 2 describes the REPA informative contents. In Section

[1] Any opinions and conclusion expressed are those of the authors and do not necessarily respect the views of the Italian national institute of statistics.
[2] https://www.istat.it/it/archivio/190748.

3, the multilevel latent class analysis is introduced and Section 4 reports the obtained results. Section 5 contains some concluding remarks.

2. The statistical register for public administrations

As introduced, REPA contains information on structural and economic variables for the subset of Italian public administrations in the S13 list. It includes different subpopulations, such as local governments, regions and autonomous provinces, ministries, constitutional bodies, social security funds, sanitary districts, etc. Each subpopulation has a specific structure and classification for its economic data. The REPA production process involves data collection, harmonization and integration, as well as data review, editing, and imputation, each tailored to the individual characteristics of each subpopulation. For more details on the contents of REPA and its production process, see (Varriale *et al.*, 2024). REPA is still under development, but the design and implementation of the register are well-advanced for the subpopulation of local governments, including municipalities, unions of municipalities, provinces, mountain communities, and metropolitan cities (Varriale *et al.*, 2021). Since our analysis focuses on the subpopulation of Italian municipalities, we will use the term REPA in the following to refer to this subpopulation and not to the whole register.

The structural variables included in REPA are identifiers and register variables, territorial and stratification variables, status of activity, date of inclusion and possible exclusion from sector S13, transformation events, number of employees. The economic variables are the result of the integration and processing (i.e. imputation) of data from administrative sources. The primary source of information is the Public Administration Database (BDAP), while the auxiliary source is the Information System on the Operations of Public Bodies (SIOPE). Economic variables include accrual and cash values, for both revenues and expenditures. The accrual data for the revenue are the assessments (E1) while the cash data are the collections in accrual (E2) and the residual accounts (E3). For expenditures, the accrual data are the commitments (S1), the cash data are the payments on accrual (S2) and the residual accounts (S3). The information for both revenues and expenditures is organized into several hierarchical levels, following the structure of the certified balance sheet that all local governments are required to publish on an annual basis to certify their primary accounting data for the previous fiscal year (Guandalini *et al.*, 2021). For each statistical unit (local government), there are 148 items for revenues - classified in titles, categories and types - and 1431 items for expenditures - classified in titles, macroaggregates, missions and programs. The highest level of aggregation of the items identified in the certified balance sheet is Titles, which is the level of aggregation used in our analyses. Table 1 shows an example of a certified balance sheet (revenue) for a single municipality. The first digit of the balance sheet code represents the title, and to derive the information at title level for each economic variable E1, E2 and E3, it is necessary to sum the economic values (symbol "xxx") identified by the same title.

Table 1 – *Example of a certified balance sheet for a single municipality. Revenue.*

Item	Balance sheet code	Title	E1	E2	E3
1	1010101	1	xxx	xxx	xxx
2	1010102	1	xxx	xxx	xxx
…	…	…	xxx	xxx	xxx
…	3059900	3	xxx	xxx	xxx
…	…	…	xxx	xxx	xxx
148	9029900	9	xxx	xxx	xxx

Concerning the revenue, in our analysis we use per capita values of: E1 - Assessments (accrual data), and E2 - Collections in accrual (cash data). We only use the information from the first 3 Titles, which sum is the current revenue: Current revenue tax based, contributory and equated (T1), Current transfers (T2), and Non-tributary revenue (T3). Concerning the expenditures, we use per capita values of: S1 - Commitments (accrual data), and S2 - Payments on accrual (cash data). We only use the information from the first 2 Titles: Current expenditures (T1), and Capital accounts expenses (T2). We consider the subpopulation of Italian municipalities included in the S13 list in 2021 and analyse their economic management behaviour, both in terms of revenue and expenditure, by applying a multilevel latent class analysis. The objective is to cluster both the municipalities and their regions to highlight some typical profiles of economic behaviour.

3. Multilevel latent class approach for REPA

Latent class (LC) analysis (Goodman, 1974) is a well know approach in social science research usually applied to clustering or constructing typologies with observed variables. With a hierarchical data structure, the usual assumption of independence of observations is violated, and a multilevel approach allows correct inferences treating the units of analysis as dependent observations (Snijders and Bosker, 2012). Multilevel data occur when there are data nested in several hierarchical levels, e.g. students in institutes, institutes in school districts, or patients in hospitals. In the case of a two-level data structure, (that is the case of our data) level 1 units are also known as lower-level or individual units, while level 2 units are higher-level or group units. The basic idea of a multilevel LC model, introduced by

Vermunt (2003) is that some of the model parameters are allowed to differ across level-2 units.

In order to describe our data we adopt a three index notation for responses: y_{kji} is the observed value on the *i-th* variable $(i = 1, \dots, I)$, on *j-th* lower unit $(j = 1, \dots, n_k)$ in the *k-th* $(k = 1, \dots K)$ higher level unit. In our analysis, according to the hierarchical nature of the data, municipalities represent lower-lever units j, and administrative regions higher-level units k. The total number of regions is *K=20*, while the number of municipalities, n_k, differs for each region. For each municipality, we observe economic variables from REPA: *6* revenues and 4 expenditures. The vectors $\boldsymbol{y}_{kj} = (y_{kj1}, \dots, y_{kji}, \dots y_{kjI})$ and $\boldsymbol{y}_k = (\boldsymbol{y}_{k1}, \dots, \boldsymbol{y}_{kj}, \dots \boldsymbol{y}_{k\boldsymbol{n}_k})$ contain the *I*-variate responses of municipality *j* from region *k*, respectively.

Applications of multilevel LC models results in simultaneously clustering individuals and groups in unobserved cluster, known as latent classes or mixtures: lower-level units are assumed to belong to one of L LCs differing in the distribution of the observed responses; higher-level units are assumed to belong to one of H higher-level LCs differing in the distribution of the lower-level LCs. The unobservable variables representing the lower- and higher-level classes membership are denoted by $x_{kj} = l$ $(l = 1, \dots, L)$ and $w_k = h$ $(h = 1, \dots; H)$, respectively.

The multilevel LC model (Vermut 2003) can be expressed in two basic equations. The first define the (mixtures) model for $f(\boldsymbol{y}_k)$, the marginal density for the full response vector of group *k*; that is:

$$f(\boldsymbol{y}_k) = \sum_{h=1}^{H} P(w_k = h) \prod_{j=1}^{n_k} f(\boldsymbol{y}_{kj} \,|w_k = h) \quad (1)$$

where $P(w_k = h)$ is the probability that group *k* belongs to higher-level LC *h*, while $f(\boldsymbol{y}_{kj}|w_k = h)$ is the conditional density for the response vector of individual *j* in group *k*. The second equation is:

$$f(\boldsymbol{y}_{kj}|w_k = k) = \sum_{l=1}^{L} P\big(x_{kj} = l|w_k = h\big) \prod_{i=1}^{I} f(y_{kji} \,|x_{kj} = l \,, w_k = h) \quad (2)$$

where $P\big(x_{kj} = l|w_k = h\big)$, is the probability that the individual *j* of group *k* belongs to LC *l* given that group *k* belongs to LC h and $f(y_{kji}|x_{kj} = l \,, w_k = h)$ is the conditional density for response variable *i* of individual *j* in group *k* given the membership to lower-level LC *l* and higher-level latent class *h*. In the first equation conditional independence between units is assumed inside each higher-level class. i.e. between municipalities. In the second conditional independence is assumed between the variables measured on the first-level units.

Finally, in order to complete the definition of the LC model, conditional densities $f(y_{kji}|x_{kj} = l\,, w_k = h)$ has to be specified, usually the choice is in the exponential family, in our case we assumed $f(y_{kji}|x_{kj} = l\,, w_k = h)$ normally distributed.

In our analysis, we applied two separate multilevel LC models for income and expenditure, the first using six variables, the second four variables. In both cases, we performed a logarithmic transformation of the original data. This choice was prompted by the observation of a strong right skewness of the distribution of all variables considered. This transformation makes it possible to emphasise the differences between the lowest values, which are the majority in both income and expenditure. The model is estimated with a maximum likelihood approach, with the program Latent GOLD, version 6.0 (Vermunt and Magidson, 2016).

A major issue in the application of the multilevel LC is the choice of the number of LCs. This issue is more complicated than in standard (single-level) LC analysis because it involves multiple, non-independent decisions (Lukociene *et al.*, 2010). To choose the number of higher- and lower-level classes we used both the information criteria (BIC, AIC), and substantive considerations related to the size and interpretation of the latent classes. In particular, following Lukociene et al. (2010), we used BIC with number of groups as the sample size, BIC(Ng), to decide on the number of higher-level classes. To contrast the tendency of information criteria to suggest a very high number of LCs, we compared the percentage difference between the information criteria of two successive models. Table 2 shows the results for income.

Table 2 – *Probability by municipal/regional cluster. Current revenue. Assessments and collections in accrual. Year 2021.*

h – l	BIC	AIC	BIC(Ng)	BIC % diff	AIC % diff	BIC(Ng) % diff
1 – 1	105549.7	105466.0	105477.9			
1 – 2	88893.8	88719.5	88744.4	-15.78	-15.9	
1 – 3	84056.0	83791.0	83828.8	-5.44	-5.6	
1 – 4	79829.8	79474.0	79524.8	-5.03	-5.2	
1 – 5	77522.9	77076.5	77140.2	-2.89	-3.0	
2 – 4	76421.2	76037.5	76092.3		-4.3	-4.32
3 – 4	75552.6	75141.1	75199.8		-1.2	-1.17
4 – 4	75352.3	74912.9	74975.7		-0.3	-0.30

4. Main results

In 2021, the number of municipalities was equal to 7,904 (S13 list). On the revenue side, we chose the solution with *H=3* latent classes (or clusters) at level-1 and *L=4* latent classes at level-2. For the expenditure, on the other hand, *H=4* and

L=2. In the following, we will use the term municipal cluster to denote the level-1 latent class and the regional cluster for the level-2 latent class. Results are presented in terms of median per capita values.

Table 3 shows the probabilities that region have of belonging to one of the 4 regional clusters, and the probability that municipalities have of belonging to one of the 3 municipal clusters, conditional to the regional cluster membership. Level-2 membership probabilities are evenly distributed, while the conditional level-1 probabilities show extreme values. Cluster 2 at regional level is characterised by a strong presence of cluster 1 at municipal level (0.78), while regional cluster 3 is characterised by a strong presence of municipal cluster 2 (0.87). In contrast, regional clusters 1 and 4 have a similar characterisation in terms of conditional probability of municipal cluster 1 (0.56 vs 0.51), but quite different for the other two municipal clusters.

Table 3 – *Probability by municipal/regional cluster. Current revenue. Assessments and collections in accrual. Year 2021.*

h	1	2	3	4
$P(w_k=h)$	0.34	0.25	0.25	0.15
$P(x_{kj}=l/w_k = h)$				
$l=1$	0.56	0.78	0.00	0.51
$l=2$	0.29	0.03	0.87	0.05
$l=3$	0.15	0.19	0.13	0.44

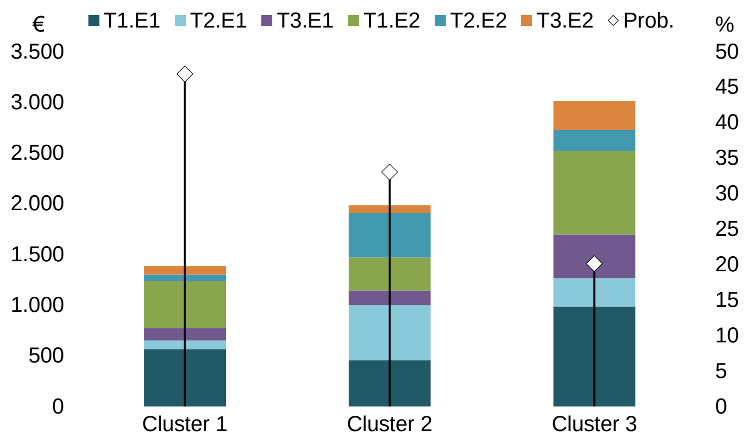
Figure 1 shows the estimated values per capita of municipal cluster. As introduced, per capita values refer to the first 3 revenue titles of municipalities (T1, T2 and T3) and refer to accrual (E1) and cash data (E2). As regards the estimated value of current revenue tax based, contributory and equated, collections (T1.E1) the maximum value was reached in cluster 3 (984 euros per capita), while the minimum values in cluster 2 (455 euros per capita). Regarding collections in accrual (T1.E2) the maximum estimated valued was equal to 823 euros per capita (cluster 3), while the minimum value was equal to 327 euros per capita (cluster 2).

The joint analysis of results in Table 1 and Figure 1 give us the possibility of describing the typical profiles of clusters, both at the municipal and regional level. In particular, for each budget item, we used the per capita value in order to make comparisons that take into account the different population of each municipality. It is possible to check whether the value of budget items for a municipality deviates excessively from the estimated values in the clusters with the highest probability and to analyse the reasons for this in order to verify the efficient use of resources.

Figure 1 – *Current revenue by municipal/regional cluster. Assessments and collections in accrual. Year 2021, per capita values in euros.*

The analysis of current revenue by municipal cluster (Figure 2) shows that the probability that a municipality belongs to cluster 1 was the highest ($P(x_{kj}=1)$ = 0.47). Cluster 1 has the per capita values of T1.E1 equal to 566 euros per capita, T2.E1 was 86 euros per capita and T3.E1 was 123 euros per capita.

Figure 2 – *Current revenue by municipal cluster and probability. Assessments and collections in accrual. Year 2021, per capita values in euros.*



It is also particularly important to be able to compare collections in accrual (E2) and Assessments (E1) by calculating the collection capacity[3] which measures the

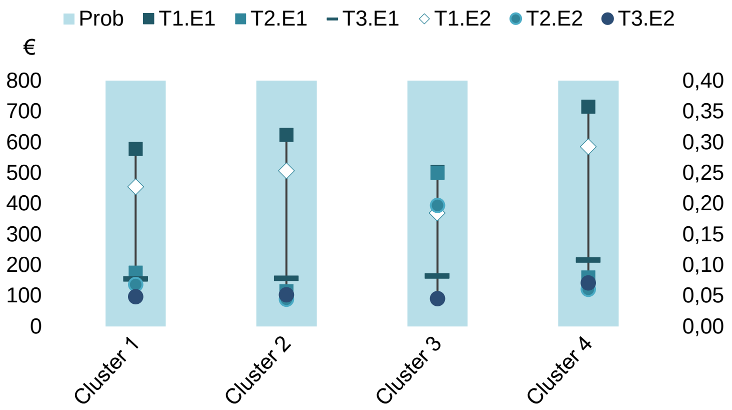
[3] The collection capacity is the percentage ratio between the collections in accrual accounts and assessments.

ability of municipalities to collect assessed revenue. As regards the titles making up current revenue, collection capacity was equal to 81% for current revenue tax based, contributory and equated, 79% for current transfers and 66% for non-tributary revenue. Comparing the values, the collection capacity for non-tributary revenue was the lowest, which means a lower efficiency in collecting that type of revenue.

Cluster analysis also allows us to calculate a number of relevant indicators used to assess the ability of municipalities to use their own revenues (tributary and non-tributary) to finance their institutional activities, in accordance with the principles laid down in the doctrine of fiscal federalism. In cluster 1 the degree of taxation autonomy[4], which measures the ability of the authority to levy resources by exercising its taxing power, was equal to 73%. The degree of financial autonomy[5], which measures the degree of autonomy of the municipality, i.e. the ratio of own revenues to current revenues, was equal to 84%.

With regard to regional clusters, the highest probability (0.34) is to belong to the cluster 1, while the lowest probability belongs to cluster 4 (Figure 3). The per capita values of cluster 1 were 578 euros (T1.E1) as regards the assessments and 455 euros (T1.E2) as regards collections in accrual. In cluster 1 the degree of taxation autonomy was equal to 64% and the degree of financial autonomy was equal to 83%.

Figure 3 – *Current revenue by regional cluster and probability. Assessments and collections in accrual. Year 2021, per capita values in euros.*



[4] The degree of taxation autonomy is the percentage ratio between current revenue tax based, contributory and equated and current revenue.

[5] The degree of financial autonomy is the percentage ratio between current revenue tax based, contributory and equated + non-tributary revenue and current revenue.

Table 4 shows estimates values of conditional probability for the current expenditures (T1) and capital accounts expenses (T2). Also for expenditure, level-2 membership probabilities - $P(w_k{=}h)$ - are evenly distributed, while the conditional level-1 probabilities - $P(x_{kj}{=}l/w_j = h)$ - show extreme values. Cluster 2 and 4 at regional level are characterised by a strong presence of cluster 1 and cluster 2 at municipal level, respectively (0.83 and 0.93), while regional cluster 1 and 3 are less extreme.

Table 4 – *Probability by municipal/regional cluster. Current expenditures and capital accounts expenses. Commitments and payments in accrual. Year 2021.*

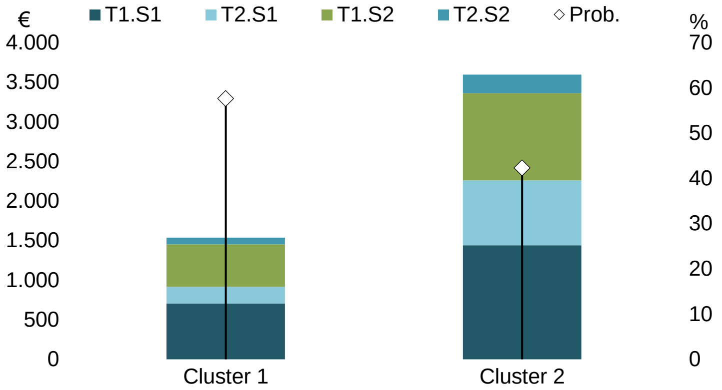
h	1	2	3	4
$P(w_k{=}h)$	0.35	0.29	0.20	0.15
$P(x_{kj}{=}l/w_j = h)$				
$l{=}1$	0.66	0.83	0.46	0.08
$l{=}2$	0.34	0.17	0.54	0.93

The per capita values of current expenditure (T1), both commitments and payments in accrual (S1 and S2), were higher in cluster 2, respectively 1,441 and 1,100 euros (Figure 4). Also capital accounts expenses (T2), commitments and payments in accrual, were higher in cluster 2 (respectively 819 and 237 euros).

Figure 4 – *Current expenditures and capital accounts expenses by municipal/regional cluster. Commitments and payments in accrual. Year 2021, per capita values in euros.*

Figure 5 shows estimated values and probability ($P(x_{kj}{=}1)$) referring to municipal cluster for expenditure. The probability for a municipality to belong to cluster 1 was the highest (0.58) but with lower per capita values. In cluster 1, which has the highest probability, the spending capacity[6], which measures the ability of municipalities to pay the amounts committed, was equal to 76% for current expenditures (T1) and 41% for capital accounts expenses (T2). In cluster 2 the spending capacity was lower for capital accounts expenses (29%), while was equal to cluster 1 for current expenditures. This means, firstly, that for capital accounts expenses the percentage of expenditure realised, compared to committed sums, was 35 percentage points lower than for current expenditure (cluster 1). Secondly, the municipalities included in cluster 1 are able to spend more of the sums allocated for long-term investments than those included in cluster 2.

Figure 5 – *Current expenditures and capital accounts expenses by municipal cluster and probability. Commitments and payments in accrual. Year 2021, per capita values in euros.*

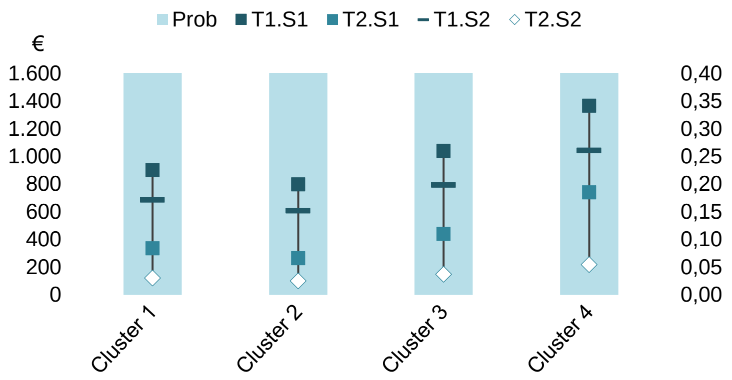


In Figure 6 the analyses of estimated values and probabilities refer to regional cluster shows that cluster 1 has the highest probability (0.35).

As regards the commitments, current expenditures was equal to 903 euros per capita and capital accounts expenses was equal to 335 euros per capita. The spending capacity was equal to 76% for current expenditures, while as regards the capital accounts expenses the indicator was equal to 36%. The same considerations made for municipal clusters regarding the ability to spend all budgeted resources also apply to regional clusters.

[6] Spending capacity is calculated as the percentage ratio between payments on accrual accounts and commitments.

Figure 6 – *Current expenditures and capital accounts expenses by regional cluster and probability. Commitments and payments in accrual. Year 2021, per capita values in euros.*



5. Concluding remarks and further developments

The statistical register for Public Administrations REPA is an object of the Italian Integrated System of Statistical Registers (ISSR), with information on structural and economic variables on a subset of the Italian PA. This paper aims to analyse the economic management behaviour of Italian municipalities. We adopt the multilevel latent class analysis which allows to simultaneously cluster municipalities and regions in groups with some typical profile. The information used for the analysis are the revenues and the expenditures. For each budget item, we used the per capita value in order to make comparisons that take into account the different population of each municipality.

The obtained results shows the characteristics of the municipal and regional cluster in terms of cluster size and value of each budget item analysed. It is also possible to calculate a number of relevant indicators used to assess the ability of municipalities to finance their institutional activities, in accordance with the principles laid down in the doctrine of fiscal federalism and to check whether for a municipality the value of the indicators and the budget items making up the clusters deviates excessively from the estimated value and to analyse the causes in order to verify the efficient use of resources. The result obtained can also help politicians and policymakers by providing useful information on the most common budget structures and the main characteristics of the main budget items.

We chose to analyse pro-capita data in order to focus on the economic structure of lower and higher units (municipalities and region) and partially absorb the effect of the different magnitude of the municipalities. On the other side using median rather than means allows a description of clusters not affected by extreme values. Further analysis could take into account other variables as dimension of municipalities, geographical collocation, administrative characteristics, both as a support to cluster description or to carry out specific conditional analysis. Moreover, comparing the obtained results with those from an analysis that considers the two levels (municipalities and region) separately could provide insights into the substantive findings offered by multilevel techniques.

References

GOODMAN L.A. 1974. The analysis of systems of qualitative variables when some of the variables are unobservable: Part I -A modified latent structure approach, *American Journal of Sociology*, Vol.79, pp. 1179-1259.

GUANDALINI A., PASSANTE D., VARRIALE R. 2022. The revenues of local governments in the statistical register for public administrations: inequality decomposition by sources, *RIEDS*, Vol. 76, No.3, pp. 17-28.

LUKOCIENE O., VARRIALE R., VERMUNT J.K. 2010 The Simultaneous Decision(s) about the Number of Lower- and Higher-Level Classes in Multilevel Latent Class Analysis, *Sociological Methodology*, Vol. 40, pp. 247-283.

LUZI O., ALLEVA G., SCANNAPIECO M., FALORSI P.D. 2019. Building the Italian Integrated System of Statistical Registers: Methodological and Architectural Solutions. *ESS Workshop on the use of administrative data and social statistics*, Valencia, 4-5 June2019.

SNIJDERS T.A.B., BOSKER R. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition*. London etc.: Sage Publishers.

VARRIALE R., LORI M., MANTEGAZZA F. 2021. Stato di avanzamento dei lavori Focus: Frame PA enti territoriali. Nota tecnica Istat, Dicembre 2021.

VARRIALE R., ALBO N., CASAGRANDE C., OLIVIERI V. 2024. The statistical register for public administrations: the reference framework and some methodological aspects. *RIEDS,* Vol. 78, No.1, pp. 173-184.

VERMUNT J.K. 2003. Multilevel Latent Class Models, *Sociological Methodology*, Vol. 33, No. 1, pp. 213-239.

VERMUNT J.K., MAGIDSON J. 2016. *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

Roberta VARRIALE, Sapienza University of Rome, roberta.varriale@uniroma1.it
Domenico PASSANTE, Istat, passante@istat.it
Cecilia VITIELLO, Sapienza University of Rome, cecilia.vitiello@uniroma1.it