rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2-3
2024

In this issue:

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2-3
2024

In this issue:

Istat

rivista di statistica ufficiale/review of official statistics
n. 1-2-3/2024

Four-monthly Journal: formerly registered at the Court of Rome, Italy (N. 339/2007 of 19th July 2007).

e-ISSN 1972-4829
p-ISSN 1828-1982

The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-anonymised approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale/Review of official statistics.

The new business statistics framework at the territorial level

Alessandro Faramondi[1], Silvia Lombardi[1], Roberto Sanzo[1]

Abstract

Recent empirical analyses of the contemporary territorial reconfiguration of productive activities have highlighted the need for official statistics to provide detailed economic information at the territorial level. The implementation of the new "Territorial Structural Business Statistics Register" (T-SBS Register) provides estimates of key economic account results at the establishment level, enabling highly granular territorial analysis. By integrating multiple data sources, including a direct survey of local units, the new T-SBS Register is fully consistent with Eurostat's Structural Business Statistics framework for enterprises. This paper presents a new methodology for estimating value added at the territorial level and offers insights into its potential for empirical economic analysis.

Keywords: Territorial Structural Business Statistics Register, data integration, value added at the territorial level, value added at the establishment level.

DOI:https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2024.01.

1 Alessandro Faramondi (faramondi@istat.it); Silvia Lombardi (lombardi@istat.it); Roberto Sanzo (sanzo@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

The growing strand of literature focused on analysing local economic development paths has so far been constrained by the availability of sub-regional economic data. In fact, employment has been the key metric of economic performance, with other measures unavailable at the micro level. Hence, territorial economic data used in pillar territorial analyses of the Italian production system, its development and transformation have so far been based on employment, with information sourced from Business Registers of Local Units, the only available official statistics source (see, among others, Iuzzolino 2004; Istat 2005; Istat 2015a and 2015b). This has forced researchers devoted to analysing territorial efficiency and growth to build several territorial databases using statistical techniques to estimate local wealth indicators based on available data and research objectives (Faramondi 2008; Bollino and Espa 2016; Bigerna et al. 2026).

In this empirical research context, the Italian National Institute of Statistics (Istat) has improved the supply of statistical information at the territorial level by intensively using administrative data (Nordbotten 2010) within the Eurostat regional Structural Business Statistics (SBS) framework. Since 2018, Istat has published a new statistical Register of economic units at the territorial level each year, known as the "Territorial Structural Business Statistics Register" (hereafter, T-SBS Register). Its provision represents a significant innovation in the business data arena for official statistics, as it increases territorial granularity compared with standard statistical domains (Menghinello et al. 2020). Official regional statistics have traditionally relied on aggregate SBS data as the primary source for territorial economic analysis. These data are based on firm-level sample surveys focused on the system of economic accounts for enterprises. As regions are a strata variable in the sample, data dissemination has so far been published at the regional level; in the case of the sub-regional domain of estimation, estimates would have been affected by sampling error in the unplanned territorial domain. Furthermore, the units of analysis are firms (legal units), without considering multi-plant firms. For these reasons, the T-SBS Register significantly improves the statistical supply of territorial information. Here, statistical units are not the enterprises but their local units, that is, their establishments and individual sites, whose information collection is mandatory within the European framework for Business Registers, in addition to enterprises and enterprise groups. In particular, statistical units of the T-SBS Register are derived from the official Business Register (BR) of Local Units (BR-LU); hence, the new Register collects, for each enterprise, economic information on its local units that contribute to the production of goods and services. For the first time, in June 2018, official statistics provided information on the main variables of the economic account for each local unit, with 2015 as the reference year.

The present work focuses on the methodological and theoretical aspects of integrating multiple data sources at the core of the new statistical Register and presents preliminary results based on new data. These results describe a new economic geography, opening the way to new scenarios for empirical analysis at the territorial level. Section 2 sets out the European framework for the official Structural Business Statistics and related data sources; Section 3 presents the genesis of the new statistical Register and its methodological aspects. Results from the new data are presented in Section 4 at the aggregate regional, municipal, and Labour Market Area (LMA) levels. Section 5 presents the quality assessment, while conclusions are drawn in the final Section.

2. The SBS system of statistical Registers

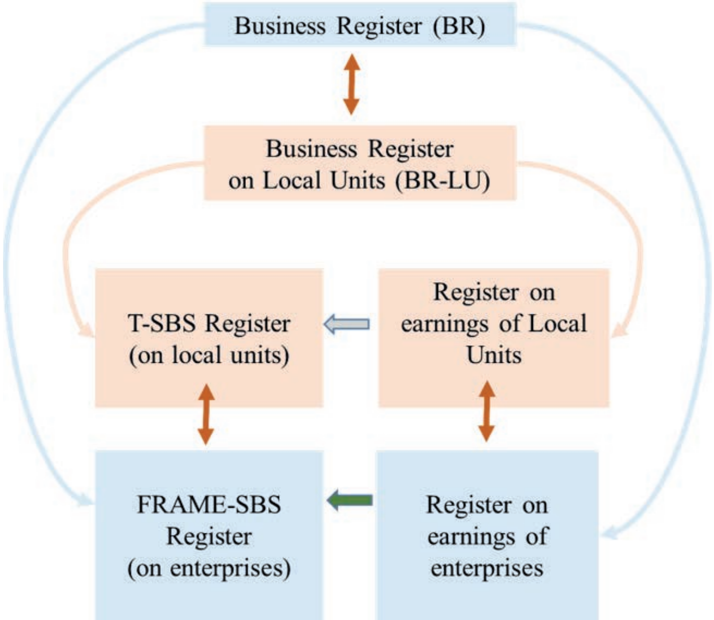
In official statistics, the Integrated System of Registers builds on the thematic development of statistical business BRs, which serve as "base statistical Registers" for economic units. Further integration with administrative and fiscal sources, as well as statistical direct surveys (both sample and total), improves topic coverage and broadens the Registers' scope within the SBS framework. Hence, integrating "basic Registers" BR and BR-LU with economic accounts information gives rise, respectively, to the SBS Register on enterprises (FRAME-SBS) and the T-SBS Register on local units. Such "thematic" registers are part of a highly integrated system, in which each Register is connected to the other components of the system, ensuring the consistency of accounting relations and the firm's productive organisation (Figure 2.1). In this framework, each enterprise-level economic variable is consistent with the sum of its local units' economic variables. As shown in Figure 2.1, the system of basic and thematic Registers of enterprises and local units can be represented in a circular form, with the outer ring representing BR and the inner ring representing the Registers of local units. In the external ring, the BR includes structural information about enterprises (specifically: sector of economic activity, number of employees, location, legal form, and group membership). It is the basis for building the extended Register on earnings, working hours, and labour costs for persons and enterprises, and, subsequently, the SBS Register on enterprises, which provides information and indicators on labour costs and the positive and negative components of the income statement, respectively.

In detail, the Register on earnings, working hours, and labour cost builds upon the integration of administrative and labour cost survey data with firm-level employment positions2. Such data integration represents the extension to the variables on labour input and labour cost for enterprises with at least one employee. The production of this Register in Istat began with the 9th Business Census in 2011, which served as the reference year. It was the result of an overall optimisation of data sources, made possible by the availability of new administrative data sources, as well as by existing sources that were usable with better timeliness. Overall, the new economic information structure

2 For details on the production process and metadata, see https://www.istat.it/scheda-qualita/registro-annuale-su-retribuzioni-ore-e-costo-del-lavoro-per-individui-e-imprese-racli/.

is composed of three levels: firms, individual workers, and their working-activity relations, and provides insights into objective aspects of working conditions (Gigante et al. 2019).

Figure 2.1 - Overview of the Integrated System of Registers on economic units



Source: Authors' processing

As far as the SBS Register on enterprises is concerned, starting from 2012 as reference year, Istat has produced the microdata containing the main economic variables, such as turnover, purchase cost of goods and services, personnel costs, value added and other more detailed variables (Luzi and Monducci 2016; Arnaldi et al. 2020). Such information is derived from an estimation process applied to data from several administrative sources, such as Chambers of Commerce, the Revenue Agency, and the National Social Security Institute. Information on companies with fewer than 100 employees is derived from a statistical treatment of administrative sources, taking into account the results of the sample survey of Small and Medium-sized

Enterprises (SMEs), which is used to construct models to impute certain variables in the SBS Register. The information relating to companies with 100 or more employees derives instead from the (census) survey on the economic accounts of large enterprises.

Finally, the internal ring focuses on local units, and its core is the base Business Register on Local Units, which provides information on the main structural characteristics of local units, such as economic activity, number of employees, and the localisation of each establishment. The territorial Register on earnings, working hours, and labour cost builds upon the BR on Local Units and, therefore, the T-SBS Register as well. These three Registers are consistent with the relevant Registers at the firm level, and all three are related to one another with respect to the economic information conveyed in the T-SBS Register. Data sources used in the three Registers are similar to those used at the firm level, with the addition that, for the definition of the BR-LU and T-SBS Register, information is also obtained from a direct survey: the coverage survey on local units (Section 3).

To ensure full consistency of the information provided in the thematic Register T-SBS Register, it is important to define its unit of analysis. In particular, in the European framework, the firm is broken down into different parts, from the local unit to the kind of activity unit (KAU)3 and its local KAU (LKAU). Currently, the statistical units of the Business Register for local units represent the best approximation of the LKau, where each elementary unit is the territorially localised production unit, and economic activity is assigned based on the prevailing economic activity criterion (i.e., the establishment). Such a definition is a proxy for LKau, since they are local units to which only one economic activity is assigned, the prevailing one. Once the basic unit closest to LKau is identified and the one represented by the BR-LU Register is taken as the territorial structure, the method for territorialising value added is defined.

3 Eurostat defines KAU as a part of an enterprise. The KAU groups together all the offices, production facilities, etc., of an enterprise that contribute to the performance of a specific economic activity defined at the class level (four digits) of the Statistical Classification of Economic Activities in the European Community (NACE Rev. 2).

3. Estimating value added at the territorial level

3.1 The methodological strategy

The estimate of value added at the territorial level builds upon the finest unit of analysis, which gives more details in the local context, that is, the local unit. As explained in the previous paragraph, the BR on Local Units contains all the local units that contribute to the production of goods and services realised by the firm itself. It provides information on their economic activity and number of employees, as well as their geographical location (for some establishments, it even includes the address, allowing geo-referencing). Hence, the statistical universe of the T-SBS Register is derived from BR on Local Units. The statistical population aligns with the scope of SBS regulation and covers NACE Rev.2 Sections B to N and Group S95 (that is, NACE Rev.1.1 Sections C to K).

The T-SBS Register is the territorial implementation of the SBS Register, which provides economic account information at the enterprise level. The innovation in the production process of territorial economic data is a methodological approach that accounts for available data sources at the local-unit level and enterprise-level information from the SBS Register. In fact, the only data source on economic information available at the territorial level is represented by the Register on earnings, working hours and labour cost on local units, which contains the information derived from the National Social Security Institute (INPS) and provides the value of the salaries of employees of the local units localised in the national territory. Therefore, wages at the local unit level are a key variable in estimating value added. Operationally, it is not possible to calculate value added as the difference between production value and its costs due to a lack of data. Therefore, the selected methodology is the "Income Approach" (Eurostat 2013). According to this method, the value added can be determined as the sum of some components per single Local Unit (LU) in the following way:

$$\widehat{VA}_{ij} = L_{ij} + \hat{K}_{ij} + \hat{R}_{ij} \quad (1)$$

where $\widehat{VA}_{ij}$ represents the value added of the j-th local unit of enterprise i, L_{ij} corresponds to the personnel costs, $\hat{K}_{ij}$ is depreciation, and $\hat{R}_{ij}$ is earnings before income and tax (EBIT), calculated in the same local unit of the same enterprise. Hence, value added is the sum of components (labour, capital, and wages) rather than the difference between production value and its costs. In fact, in calculating GDP under the Eurostat Framework of National Accounts, the income-side approach shows how value is distributed among participants in the production process.

In addition, since 2017, the reference date and depreciation value have been sourced directly from the direct coverage survey of local units (see Section 3.2). A subset of large enterprises reports the amount of depreciation for each local unit of large multi-plant enterprises to the T-SBS Register, while the remaining subset is estimated (see Section 3.3). The income approach allows, therefore, a better exploitation of statistical sources at the local unit level, thus reducing the estimation of only one variable: EBIT ($\tilde{R}_{ij}$) and depreciation and amortisation ($\tilde{K}_{ij}$) that are not provided from a direct survey.

3.2 The direct survey on multi-plant enterprises

Previous paragraphs have shown that estimates of economic variables primarily concern multi-plant enterprises, i.e., enterprises with more than one local unit, which may be located outside the planned territorial domains of the SBS framework (regions). Moreover, while administrative data integration provides information on employees and wages for local units, data on their fixed investments are not available, even though they are crucial for the correct calculation of value added. Therefore, the central role of depreciation in estimating territorial value added for local units of multi-plant firms is evident. In the Italian productive landscape, local units of multi-plant enterprises represent an important component of the business system: in 2015 (the first reference year of the T-SBS Register), there were more than 510,000 such units, representing 11.2% of the total number of local units and 35.3% of total employees, but accounting for about half of the value added (48.8%) and depreciation (55%).

The T-SBS Register sources depreciation data for establishments from a direct survey and a yearly coverage survey of firms instrumental to the production of BR-LU. Since the production of the T-SBS Register, firms are asked for information not only on the structural aspects of their establishments – such as persons employed, address, and state of activity, as well as the type of local unit - but also on depreciation, which therefore does not have to be estimated. This improvement in the direct sourcing of depreciation is relevant given the sample size and the structural composition of Italian enterprises.

The survey sample is mixed in nature, as it is based on a census of very large enterprises, both single- and multi-plant (with more than 100 persons employed), while smaller enterprises are rotated every two to three years. Overall, the sample survey accounts for about 10,000 enterprises, which cover almost one-third of local units of multi-plant firms each year. For this reason, it is not possible to provide information on depreciations for all multi-plant firms every year, as they are excluded from the direct survey. For such firms, an estimation approach is necessary (see Section 3.3).

Editing and imputation of information on collected depreciation use administrative data at the firm level to check the total value. However, the most important advantage of the direct survey is that firms are asked to estimate depreciation on fixed investments even when they lack accounting information. This case can be quite frequent for smaller enterprises, as it is not mandatory information. For this reason, the data editing steps are guided by the principle of retaining the most accurate survey data, while incomplete information (depreciations provided for some local units only, without explanation) is not taken into consideration and is therefore estimated in the following phase. Therefore, missing values and measurement errors are first checked in detail. Secondly, consistency checks are performed between the administrative value of depreciation at the firm level and the sum of surveyed establishment values. Thirdly, quality evaluations are devoted to cases of concentrated depreciation (i.e., enterprises with depreciation in only one establishment). Finally, the analysis of correlations between depreciation and persons employed by the domain across both enterprises and establishments concludes the process. Overall, since the introduction of the economic section in the questionnaire, the use of survey data has increased the number of firms whose depreciation was collected directly, rising from 13% in the first survey (reference year 2017) to 16% in the following year.

3.3 The estimation approach

As previously mentioned, the only addendum to equation (1) provided by a statistical source is a proxy for the cost of personnel, in this case, the amount of wages. The Register on earnings, working hours, and labour costs of local units is therefore the source of input L_{ij}. With respect to the remaining variables, their estimation at the local unit level takes place in a sequential manner by collecting information on depreciation first ($\hat{K}_{ij}$), and then by exploiting the former variable to estimate EBIT ($\hat{R}_{ij}$).

3.3.1 Estimation of depreciation and the hypothesis on multi-plant enterprises

Depreciations per local unit are collected annually through a direct survey (see Section 3.2 above). However, depreciation estimates are made for the sub-population of large enterprises that do not provide such information. To move forward with the value-added estimation procedure, it is necessary to identify the group of companies that best represents the performance of the local units and, on this basis, determine the best model for estimating depreciation at the territorial level. In this regard, all economic information on single-plant enterprises is derived from the SBS Register. Such firms are, by definition, composed of a single establishment that coincides with the legal unit and carries out only one predominant activity. In addition, large enterprises that carry out only one prevalent activity ("single-function" large enterprises) are included in this group[4]. The basic hypothesis concerns the local units of multi-plant enterprises, the only ones that need to be estimated at the plant level in economic accounts. They are supposed to have an "economic behaviour" completely similar to that of "single-function" large enterprises and single-plant enterprises that operate in the same sector. For this reason, great emphasis has been placed on the variable relative to the NACE classification, which has been kept as detailed as possible across the estimation domains to better identify the underlying model.

In initial attempts to estimate, Faramondi et al. (2018) describe the methods. The fundamental choice was the logic of identifying the domains of estimation in which to estimate depreciation: we started from a very detailed situation considering the 5-digit NACE, 10 classes of employees (0, 1, 2-3,

4 The "multi-functional" large enterprises are not included, as they could have made spurious the existing relations between the variables that would then be used in the model.

4-5, 6-9, 10-19, 20-49, 50-99, 100-249, 250+) and the Italian geographical areas (North, Centre, South and Islands), and then verify whether there were a sufficient number of units within the cell thus identified. In the event that this did not happen, the domains were collapsed, initially eliminating the geographical distribution, which was not considered particularly significant for the estimation of depreciation in local units, and then aggregating the classes of employees, only as a last resort, which was considered a less detailed level of the NACE classification. Since depreciation is a variable that is not always present and linked to other variables in the profit and loss account, modelling the interrelations between it and the other variables available at the territorial level may be difficult and even uncertain. Among the various studies carried out during the preparation of the activities for estimating depreciation, it was decided to use the average value of depreciation per employee calculated in a very fine domain (starting from the 5-digit NACE) for single-plant and unified enterprises; the availability of information on wages from the source RACLI also suggested to consider separately two sub-sets, that of enterprises with employees and that of enterprises without employees. The solution used was therefore certainly very simple, even if its use has guaranteed a set of "prudent" estimates, not having more certain reference points with regard to a very great territorial detail: it is desirable, however, that we can, in the near future, deepen other methods (regressive models, neural networks, etc.) in order to ensure the best possible result.

Depreciations are calculated using an average indicator for each domain (^{d}) derived from single-plant enterprises. Such an indicator is:

$$\widehat{K}_{ij}{}^{d} = K_i / Emp_{ij}$$

Domains are combinations of the four-digit NACE codes (classes of economic activity), size (classified by employees), and macro-geographical area[5].

3.3.2 *Estimation of Earnings Before Interest, Taxes, Depreciation, and Amortisation (EBITDA)*

To estimate the last addend of the income approach equation (1), namely EBIT, we need to introduce an additional hypothesis. We assume that the

5 Scarcely populated domains are collapsed into higher hierarchical level domains. In case of zero employees in the local unit object of estimate, the average domain indicator d is multiplied conventionally by 0.5.

incidence of EBITDA in the local unit equals that of the sum of wages and depreciation in the unit itself.

Therefore, remembering that for each local unit "j" of the enterprise "i" we have

$$\widehat{VA}_{ij} = L_{ij} + \hat{K}_{ij} + \hat{R}_{ij} \qquad (1)$$

Where $\widehat{VA}_{ij}$=value added, L_{ij}=salaries, $\hat{K}_{ij}$= the value of depreciation and amortisation, and $\hat{R}_{ij}$=EBITDA. As $\hat{K}_{ij}$ is calculated in the previous phases, by using the Register on earnings, working hours and labour cost of local units, it follows that

$$\widehat{VA}_{ij} - \hat{R}_{ij} = L_{ij} + \hat{K}_{ij} = \hat{X}_{ij} \qquad (2)$$

Calculating on the SBS Frame at the enterprise level, the same amount $X_i = L_i + K_i$, we can determine the EBITDA per local unit as follows

$$\hat{R}_{ij} = \frac{\hat{X}_{ij}}{X_i} R_i \qquad (3)$$

Finally, by replacing in (1) the value of $\hat{R}_{ij}$ we obtain an estimate of the value added of the local unit *j* of the firm *i* by summation of the addends.

3.3.3 Estimation of other variables

To estimate the other variables of the profit and loss account for local units, the methodology of the SBS Register is followed. Therefore, estimates are carried out in three phases:

1. Estimation of the positive components of the value of production and changes in stocks (VP) and the cost components (C) per local unit, starting from the estimated value added, using the formula already used in the SBS Register, namely Value added = VP - C;
2. Estimate of the components of VP (revenues from sales and services, increases in fixed assets, other revenues, changes in contract work in progress and changes in inventories) and of C (purchases of goods and services, use of third-party assets and other operating costs).
3. Quadrature of variables at the enterprise and local-unit levels.

To estimate VP and C, we selected Predictive Mean Matching (PMM) (Rubin 1986; Little 1988). A statistical model defines a distance function under which the donor provides the value closest to the recipient's "predicted average". The variables used to calculate this distance were value added and salaries for local units of enterprises with employees, while value added was the only variable for those without employees. The imputation domain was set at a 5-digit NACE code; however, if the number of donors was fewer than 100 or the ratio of recipients to donors was less than 50%, the domains were aggregated.

All donors were identified using the same method as for estimating value added: only single-plant and unified companies were included; obviously, all receivers were identified by all the local units of the multi-plant firms. The application of the method has therefore allowed us to impute the values of VP and C in a way that is completely consistent with the previously estimated value added per local unit. However, the method does not guarantee that the sum per enterprise of the two variables estimated at the local unit level matches the overall enterprise data; therefore, a first redistribution of the values attributed to VP and C is necessary to make their sum equal (except for approximations) to the total enterprise, while at the same time ensuring that VP-C still provides the value added.

The Nearest Neighbour Donor (NND) method (Kalton and Kasprzyk 1982; Chen and Shao 2000; Luzi et al. 2007) was used to impute the component variables of VP and C. In this method, which is a generalisation of the previous method, missing or inconsistent values detected in a receiving unit are imputed using values from a donor considered the "closest" based on a distance measure derived from a vector of observed, auxiliary, or matching variables. The distance function used is Euclidean, and, as matching variables, VPs, Cs, and Register on earnings, working hours, and labour cost for local units have been used; for enterprises with employees, only VPs and Cs are used; for those without employees, only Register.

What is actually imputed in this case is not the value of the variable itself, as it was previously done, but a series of ratios of the variables to be imputed with respect to VP or C: the imputation of these ratios derived from the donor, applied then to the total VP or C estimated per local unit, provides the value of the attributed variable. For some sectors (supermarkets and airlines) and for one firm in particular (Post Office), the imputation method described above was not used because of the difficulty of finding single-plant, unified companies that could serve as donors, given the specificity of these companies.

In these cases, the firm's economic value was therefore distributed among the local units based on their weight in the total, measured by employees and/or value added.

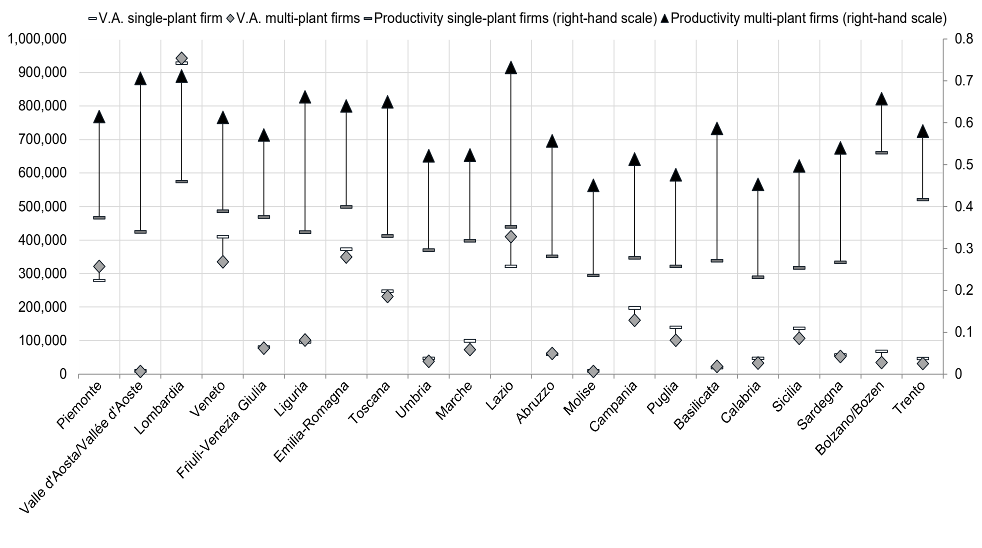
The use of the NND method, as constructed, guarantees congruence between the VP and the sum of its components, and between the C and the sum of its components. As before, however, the method does not guarantee that the sum per local unit of the various components, taken separately, equals the enterprise-level data for the same variable. For this reason, it was necessary to move on to a quadrature phase, which was also quite complex, as it had to reconcile the values of all variables estimated per local unit with those at the firm level, ensuring their internal and overall consistency with the data provided by the SBS Frame.

4. Quality assessment on value-added estimates of multi-plant firms

The Sections above have shown that estimating economic variables primarily concerns multi-plant enterprises. This Section focuses on the analysis of F-SBS estimates, without entering into comparisons with other sources, with particular attention to selected variables and indicators used to assess the quality of the estimated output.

Local units of multi-plant enterprises represent an important component of the business system. As for single-plant firms, they are mainly localised in the Northern and Central regions, while their employees are concentrated in a few regions: Lombardia, Piemonte, Veneto, Lazio, and Emilia-Romagna, meaning that their average size is higher in these regions (Figure 4.1). Despite the uneven distribution of multi-plant firms' local units and employees, their aggregate value added at the regional level equals that produced by single-plant firms, thereby generating differentials in nominal labour productivity, measured as value added per employee. Three stages feature this part of the work.

Figure 4.1 - Value added and productivity of local units of single and multi-plant enterprises, total economy. Year 2015 (Values in thousands of euros)



Source: Authors' processing of the T-SBS Register data

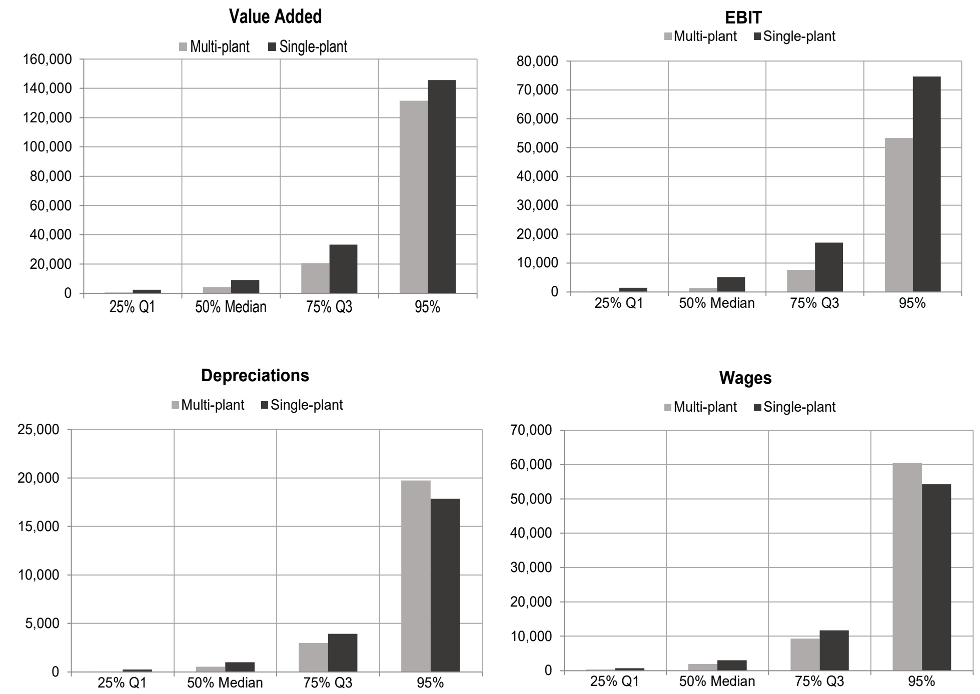
In the first stage of the analysis of the estimate of value added at the national level, we concentrated on the study of the variability of the distributions of the economic variables at the level of local units in each domain, that is, combinations of regions and divisions of economic activities according to the NACE 2007 classification. Economic results for local units of multi-plant and single-plant enterprises were investigated separately as distinct subsets of the local unit population. Overall, the investigated domains are 1,508, and the variables are value added, amortisation, earnings before income and tax (EBITDA), turnover, and their relative terms (per employee). In order to compare variability among statistical distributions of such variables, we selected the coefficient of variation as a metric, since it allows the comparison of the dispersion of data with different ranges of variation:

$$CV = \left(\frac{\sigma}{|\mu|}\right) * 100 \quad \mu \neq 0$$

Where σ is the standard deviation and |μ| is the absolute value of the mean of each distribution. Marking IQR = (Q3-Q1), the interquartile difference, the lower adjacent value is the smaller value among the observations that is greater than or equal to Q1-1,5*IQR. The upper adjacent value is the largest value among the observations that is less than or equal to Q3+1.5*IQR. Therefore, the CV contained between Q1-1,5*IQR and Q3+1,5*IQR is considered acceptable. As far as multi-plant firms are concerned, 97.5% of the coefficients of variation are acceptable, as they fall within the range defined by the adjacent lower and upper extremes of their respective distributions. The number of unacceptable CVs of multi-plant firms is consistent with that of single-plant firms.

In the second stage, a municipality-level analysis of local units of multi-plant companies focuses on 7,822 municipalities with local units (97.3% of the total) in 2015. The comparison of variables between local units of single- and multi-plant firms shows wide differences in their distributions. Graphically (Figure 4.2), the value of IQR of the municipal aggregates of local units of multi-plant firms is systematically lower than that of the localised ones. The comparison between the distributions of value added, EBITDA, amortisations, and wages of the two subsets of local units shows similarities between amortisations and salaries and wages, with the 95th percentile of the localised multiples significantly higher than the localised ones (third and fourth quadrants).

Figure 4.2 - Distribution of value, EBIT, depreciation, wages and salaries of local units of single- and multi-plant enterprises at the municipal level. Year 2015
(Values in thousands of euros)



Source: Authors' processing of the T-SBS Register data

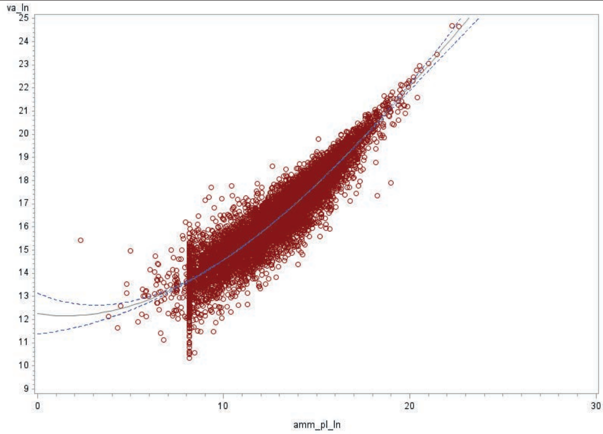
Lastly, in the third stage, specific attention is given to evaluating the contribution of local units of multi-plant firms to the formation of aggregate value added at the municipal level. By relating the total value added generated by total local units as a function of the amortisations generated only by local units of multi-plant firms at the municipal level, a high correlation is shown, as expressed by the Pearson coefficient equal to 0.97683.

The relation examined is:

[Value Added (multi-plant + single-plant), Municipality] = f [(Amortisations multi-plant, Municipality)]

The presence of high heteroscedasticity requires a logarithmic transformation of the variables, which allows the non-linearity of the phenomenon to be represented in Figure 4.3 by a third-degree polynomial relation. The non-proportional contribution of the amortisation of localised multiples indicates that aggregated returns to scale are not constant.

Figure 4.3 - Plot of depreciation of local units of multi-plant enterprises (x) and total value added (y). Year 2015



Source: Authors' processing of the T-SBS Register data

5. Results from the new data: a breakthrough in regional studies

The availability of the first extended Register of economic variables at the territorial level increases the accuracy of official territorial economic information.

The accessibility of accurate data at the establishment level enables in-depth investigations using a variety of sub-regional geographies as units of analysis, either functional or census-based. In fact, the dissemination of results is organised by structural dimensions and subnational geographies. These geographies make use of both administrative areas of the Italian territory (i.e., regions, Autonomous Provinces, metropolitan areas, municipalities, and their subset of capital municipalities of provinces) as well as official functional geographic units of analysis, such as Labour Market Areas (LMAs) and their official classifications (Istat 2015c; Faramondi et al. 2022). However, the availability of microdata enables the finest aggregation, which is instrumental to research interests in regional studies.

This Section presents such an increase at the regional level (Section 5.1), which is the administrative division compliant to SBS Regulation, at the municipality level (Section 5.2), that is, the fines administrative division, and at the intermediate territorial level (Section 5.3) in terms of functional partitions of geographies.

5.1 Measuring the regional value added of Southern Italy

Its comparison with the information disseminated by Istat at the regional level to comply with EU Regulation N. 295/2008 within the SBS framework "T-SBS data" reveals an average discrepancy of 0.1% in value added, as well as a territorial aggregation of uncovered value added in the Southern regions of Sardegna and Calabria (Table 5.1).

In terms of geographical distribution, indeed, the new data highlight an increase in value added in the South by 1.5%, in the Centre by 1.1%, and in the North-East by 0.2%, and a decrease in the North-West (-1.5%). At the regional level, the estimate of value added for 2015 shows significant differences across more than half of the Italian regions. In particular, among the northern regions, Lombardia, which contributes more than 26% to national value added, records a 2.3% decrease in value added (-4,493,997 thousand euros), while Friuli-Venezia

Giulia shows a 591,413 thousand euros increase (+3.9%). Across the regions of Central and Southern Italy, with the exception of Basilicata, which recorded a 0.9% decrease, there was a general increase in value added. The new estimate shows, in fact, a significant increase in absolute terms for Lazio and Tuscany (respectively +1,018,424 thousand euros and +1,006,987 thousand euros), and an increase of 7.8% for Sardegna, amounting to 791,791 thousand euros.

Lombardia remains the region that holds the first place in the ranking, achieving over 187 billion euros of value added, followed by Veneto (74.5 billion), Lazio (73.1 billion) and Emilia Romagna (72.1 billion), while the last places are held by Valle d'Aosta/*Vallée d'Aoste* (1.7 billion), Molise (1.7 billion) and Basilicata (4.2 billion).

Table 5.1 - Comparison of value added by Italian regions. Year 2015 (thousands of euros and percentages)

REGION	T-SBS	SBS	Difference %	
	Value added (VA1)	Value added (VA2)	Value added	*Value added per employee*
Piemonte	59,985,432	59,453,050	0.9	-0.4
Valle d'Aosta/*Vallée d'Aoste*	1,698,926	1,772,874	-4.2	0.7
Lombardia	187,004,426	191,498,423	-2.3	-0.7
Bolzano/*Bozen*	10,166,212	10,474,455	-2.9	-0.4
Trento	7,712,177	7,821,357	-1.4	-3.6
Veneto	74,479,647	74,570,865	-0.1	0.2
Friuli-Venezia Giulia	15,724,130	15,132,717	3.9	1.4
Liguria	19,690,665	19,459,277	1.2	0.0
Emilia-Romagna	72,150,588	71,732,719	0.6	0.1
Toscana	47,883,708	46,876,721	2.1	0.5
Umbria	8,457,732	8,376,893	1,0	-0.5
Marche	17,165,792	17,113,342	0.3	0.1
Lazio	73,129,182	72,110,758	1.4	2.1
Abruzzo	12,011,393	11,914,397	0.8	-0.4
Molise	1,729,221	1,699,016	1.8	3.2
Campania	35,696,255	35,634,715	0.2	0.5
Puglia	23,932,108	23,517,620	1.8	0.3
Basilicata	4,199,596	4,236,826	-0.9	-2.0
Calabria	7,937,560	7,751,935	2.4	-0.2
Sicilia	24,262,515	24,022,324	1,0	0.1
Sardegna	10,896,845	10,105,054	7.8	7.4
Geographical areas				
North-West	268,379,449	272,183,624	-1.5	-0.6
North-East	180,232,753	179,732,113	0.2	0.0
Centre	146,636,414	144,477,714	1.1	1.2
South	120,665,493	118,881,887	1.5	0.8
ITALY	**715,914,109**	**715,275,338**	**0.1**	**0.1**

Source: Authors' processing of the Istat T-SBS Register data

The new information base also shows that Northern regions, with about 2.3 million local units, account for more than 60% of national value added. In the regions of the Centre reside 21.5% of the local units, which contribute 20.5% to national value added, while the South accounts for 16.9% (27.8% of local units). In 2015, data on labour productivity across territorial divisions confirm higher values in the North Western regions (73.6 thousand euros per employee), followed by the North Eastern regions (67.7 thousand euros per employee), the Centre (65.2 thousand euros per employee), and the South (50.8 thousand euros per employee).

5.2 The first official estimate of value added at the municipal level

The T-SBS Register has also filled a significant information gap at the municipal level. The municipal ranking by value added shows only capital cities in the top twenty positions, which together represent 40% of national value added, produced by 36.3% of the local units of the national territory. Specifically, Milano ranks first, with a value added of more than 52 billion euros, accounting for 7.3% of national value added. Roma ranks second with 51,8 billion, equal to 7.2% of the total, due to the considerable weight of the value added generated by construction and utilities. Torino, Genova and Napoli follow, generating wealth equal to 2.1%, 1.5% and 1.3% of the Italian value added, respectively.

Overall, the first five municipalities in the ranking, which account for 14.1% of local units, contribute 19.5% to national wealth. The last positions, by contrast, are mainly held by municipalities in the Southern regions, particularly Sardegna (with seven capitals in the last seven positions). The Centre ranks at the bottom of the list, with the capitals of Fermo and Rieti.

The analysis by macro-sector highlights the primacy of the Municipality of Roma in industry, where the value added, generated by 3.0% of the national economic units, exceeds 10 billion euros, while Milano stands at just under 9 billion euros (Table 5.2), In detail, the comparison between the two capitals shows the prevalence of manufacturing only for the Municipality of Milano (+24% of value added compared to Roma) against a lower contribution in the sectors of energy and water supply (-70.5%) and construction (-8.7%). In terms of services, the top five positions confirm the ranking already observed on the total economy; it is worth noting the rise, among the municipalities of the South, of Palermo, in tenth place with 3,9 billion euros of value added, and Bari, in twelfth place with 2,9 billion euros.

Table 5.2 - Top 20 capital municipalities in the ranking of value added in industry and services. Year 2015 (thousands of euros and percentage values)

Industry					Services				
		Value added		% Local units in Italy			Value added		% Local units in Italy
Ranking	Municipality	Absolute values (in thousands of euros)	% In Italian value added		Ranking	Municipality	Absolute values (in thousands of euros)	% In Italian value added	
1	Roma	10,780,374	3.58	3.03	1	Milano	43,533,403	10.49	4.62
2	Milano	8,973,454	2.98	2.13	2	Roma	41,020,448	9.88	6.33
3	Torino	3,741,023	1.24	1.27	3	Torino	11,469,419	2.76	1.97
4	Genova	2,900,855	0.96	0.78	4	Genova	7,598,369	1.83	1.14
5	Napoli	2,038,774	0.68	0.78	5	Napoli	7,346,710	1.77	1.72
6	Parma	1,995,463	0.66	0.33	6	Firenze	5,919,073	1.43	1.06
7	Firenze	1,859,090	0.62	0.63	7	Bologna	5,771,064	1.39	1.03
8	Bologna	1,839,893	0.61	0.46	8	Venezia	4,267,460	1.03	0.59
9	Modena	1,757,548	0.8	0.33	9	Verona	3,925,419	0.95	0.59
10	Brescia	1,475,197	0.49	0.29	10	Palermo	3,885,426	0.94	0.93
11	Venezia	1,384,874	0.46	0.34	11	Padova	3,580,466	0.86	0.63
12	Reggio nell’Emilia	1,286,896	0.43	0.42	12	Bari	2,875,989	0.69	0.64
13	Prato	1,098,139	0.37	0.78	13	Brescia	2,779,153	0.67	0.56
14	Bolzano/*Bozen*	1,092,651	0.36	0.22	14	Modena	2,261,385	0.54	0.42
15	Ravenna	1,033,683	0.34	0.24	15	Parma	2,142,369	0.52	0.43
16	Verona	995,549	0.33	0.34	16	Bergamo	2,094,611	0.50	0.40
17	Palermo	980,377	0.33	0.42	17	Reggio nell’Emilia	2,065,402	0.50	0.34
18	Padova	896,045	0.30	0.27	18	Catania	2,024,885	0.49	0.54
19	Bari	865,396	0.29	0.29	19	Bolzano/*Bozen*	1,857,650	0.45	0.26
20	Forlì	852,646	0.28	0.23	20	Cagliari	1,799,192	0.43	0.43
	ITALY	**300,812,447**	**100.00**	**100.00**		**ITALY**	**415,101,662**	**100.00**	**100.00**

Source: Authors’ processing of the Istat T-SBS Register data

In terms of labour productivity, shifts feature at the top of the ranking. With the exception of the first position, where the Municipality of Milano is confirmed, with productivity that exceeds the Italian average by one and a half times, the Municipality of Bolzano/*Bozen*, in second position, presents an equally considerable result (index number 151.3, with value added per employee of 68,9 thousand euros); followed by the Municipality of Siena (index number 132.0) and the Municipality of Brindisi, the only municipality in the South of Italy to rank first with a productivity value of 58,2 thousand (index number 127.7), before Roma, which stood at 57,1 thousand euros. At the bottom of the ranking are the municipalities of the South and the Islands, as well as some capitals of the Centre (Fermo, Rieti, Viterbo, Grosseto, Prato, Macerata, Pistoia).

In terms of macroeconomic sectors, the performance of industry in the Municipality of Bolzano/*Bozen* stands out, with nominal labour productivity of 122,9 thousand euros, more than double the Italian average, followed by the Municipality of Pavia, which ranks second. It should also be noted that Brindisi's positive result is driven by the industrial sector, in particular the energy supply sector; it is the same sector that places the Municipality of Avellino in fifth position, between Milano and Roma. In the services sector, after Milano, which stands out with 67,2 thousand euros of value added per employee, the Municipality of Siena is positioned, characterised by productivity (55,8 thousand euros) mainly related to business services. The first capital of the South appears only at the 28th position: it is Naples, with an indicator value of 39,3 thousand euros, slightly lower than the national average (index number 99.3).

Looking at the totality of the municipalities, beyond the imbalance in the distribution of labour productivity values in favour of the northern area of the country, it is possible to identify some areas of particular interest (Table 5.3). In the Centre, the Municipality of Pomezia, characterised by a strong concentration of production activities linked to large multinational groups, stands out for a significant economic result in terms of both value added (1,8 billion) and productivity (58,2 thousand); equally significant is the figure for the Municipality of Fiumicino, which, especially by virtue of the presence of the airport, recorded 1,7 billion of value added with a productivity of almost 48 thousand euros. In the southern area, the municipalities of Melfi and Foggia have a pole of industrial units. In particular, as far as the manufacturing industry sector only, the performance of the Municipality of Melfi emerges, which stands out not only among those of the South: in the ranking of non-capital municipalities, it ranks second, with a value added of 863 million and a productivity of 92,4 thousand euros, behind the Municipality of Maranello, which is close to one billion in terms of value added (962 million) and has a nominal productivity value of more than three and a half times that of Italy (213,5 thousand). The South is also represented, in the first twenty positions, by the municipalities of Atessa (671 million of value added), Modugno (543 million) and Pomigliano d'Arco (515 million), where a large part of the value added is generated by the plants of large companies operating in the sectors of the manufacture of motor vehicles and other means of transport.

Pomigliano, in particular, differs in that its average labour productivity is lower than the national figure (an index number of 82.8). In Lazio, for the municipalities of Pomezia and Aprilia, the largest contribution to value added comes from the pharmaceutical sector. Tuscany is represented by the municipalities of Sesto Fiorentino and Scandicci, both supported by the economic results of the leather goods manufacturing sector. Of particular importance, finally, is the figure for the Municipality of Vado Ligure, which is characterised by a greater impact on the chemical sector's value added and, in terms of productivity, reaches a value 5 times higher than the national one (316,2 thousand).

Table 5.3 - Top 20 non-capital municipalities by value added in the manufacturing industry. Year 2015 (thousands of euros and percentage values)

Position in the overall ranking	Position in the ranking of non-capital municipalities	Municipality	Value added (in thousands of euros)	Value added per employee	
				Absolute values (in thousands of euros)	Index Italy number =100
11	1	Maranello	962,000	213,547	360.6
12	2	Melfi	863,225	92,402	156.0
14	3	Pomezia	822,561	94,892	160.2
16	4	Fiorano Modenese	779,615	84,669	143.0
17	5	Sesto San Giovanni	685,685	126,660	213.9
18	6	Sesto Fiorentino	673,880	100,270	169.3
19	7	Atessa	670,539	62,494	105.5
21	8	Aprilia	638,615	91,782	155.0
25	9	Agrate Brianza	598,921	81,189	137.1
26	10	Scandicci	564,900	62,088	104.8
27	11	Alba	554,937	103,769	175.2
29	12	Modugno	542,827	62,837	106.1
32	13	Pomigliano d'Arco	514,725	49,018	82.8
34	14	Schio	497,857	69,468	117.3
36	15	Imola	489,571	74,029	125.0
37	16	Cesena	485,966	60,355	101.9
39	17	Arzignano	471,082	65,808	111.1
41	18	Rivalta di Torino	443,914	115,879	195.7
43	19	Vado Ligure	440,581	316,205	534.0
44	20	Fabriano	439,914	70,710	119.4
		ITALY	**213,458,400**	**59,218**	**100.0**

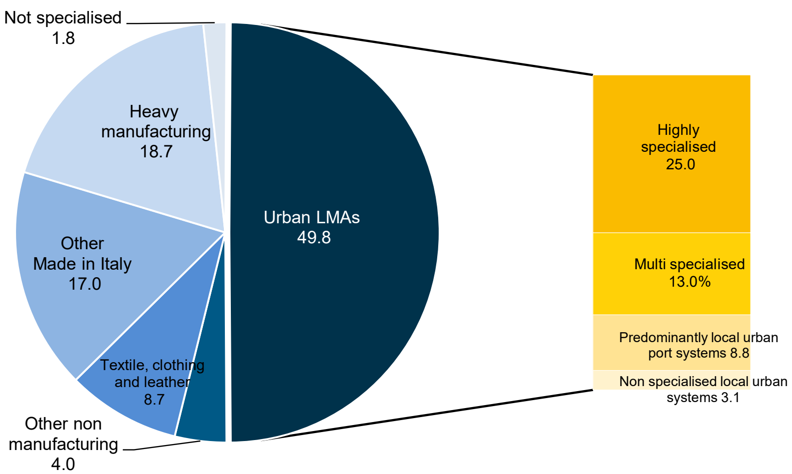
Source: Authors' processing of the Istat T-SBS Register data

5.3 Labour market areas (LMAs) analysis

Traditionally, regional sciences have made extensive use of LMAs as a unit of analysis, although, as noted above, employment has traditionally been the sole economic performance metric (Sforzi 2009; Lombardi and Sforzi 2016).

As for value added, the T-SBS Register provides interesting figures on the deep transformation underway in the Italian production system. In 2015, urban LMAs generated almost half (49,8%) of national value added in industry and non-financial services (Figure 5.1). In this category, the share of local units and employment is about 46%. Urban local systems with high specialisation play a crucial role in the national economy, realising 25% of overall value added (and 43% of service value added), followed by pluri-specialised ones (13%). At the same time, there is a strong polarisation around the LMAs of Milano and Roma. In line with the OECD (2020), the correlation between urbanisation and economic growth is evident in Italian metropolitan areas, where the economic role of commuting zones is crucial. Figure 5.1 shows that Made in Italy LMAs account for over a fourth of the national economy (textile, clothing, and leather at 8,7% and other Made in Italy specialisations at 17%), while heavy manufacturing accounts for less than a fifth (18,7%). In the former group, a significant share is covered by industrial districts, especially in the machinery, textile and clothing sectors. Finally, Made in Italy LMAs account for one-third of industrial value added and one-fifth of services, thus highlighting the territorial pattern of integration between manufacturing and services. Among Heavy Manufacturing LMAs, those specialised in the petrochemical and pharmaceutical industries realise 6,6% of overall value added.

Figure 5.1 - Value added by type of Labour Market Area. Year 2015 (percentage values)



Source: Authors' processing of the Istat T-SBS Register data

The availability of the T-SBS Register has revealed the growing developmental role of specific service sets by clarifying the relationship between service localisation patterns and the economic performance of LMAs. In particular, it highlights the role of a class of advanced services known as knowledge-intensive business services (KIBS) in empowering local manufacturing systems (Vendrell-Herrero and Wilson 2017; Lafuente et al. 2017). In 2015, KIBS employment accounted for 18,6% of overall national employment (Table 5.4). Overall, high-specialisation and pluri-specialised local urban systems have KIBS location quotients (LQ) greater than 1, indicating that employment in these LMAs exceeds the national average. This is evidence of local systems specialised in KIBS. An interesting result concerns another set of LMAs specialised in KIBS: the transport industry LMAs. However, all the Made in Italy LMAs, with the exception of those specialised in Jewels, glasses and musical instruments, have KIBS employment below the national average. This may hamper traditional manufacturing's ability to enter servitisation trajectories.

In terms of KIBS value-added generation, KIBS value-added is concentrated in urban specialised areas, which generate two-thirds of total KIBS value-added, though specific manufacturing systems also contribute. Among the above-mentioned specialisations (e.g., transport industries), the textile and clothing industries and the agro-food industries also exhibit embedded KIBS, generating a high level of value added. The proportion of KIBS value added to service activity value added offers some insight into the structural transformation of these LMAs. Figures show that the higher share is in local urban systems with high specialisation and in manufacturing LMAs specialised in transport industries, where KIBS also have the highest weight on total service value added. Moreover, Made in Italy LMAs, where KIBS employment accounts for a percentage of the total economy (14, 5%) lower than the national average (18, 6%), in fact generate increasing shares of service value added. This provides insight into how servitisation processes affect the Italian transformation process. Here, about 30% of service value added is generated by KIBS. This investigation allows the identification of the impact of different types of services on the structural redefinition of value creation. Moreover, even if urban areas seem to be a more favourable environment for KIBS location, specialisation is important for KIBS value creation. Finally, in line with recent servitisation literature (Vendrell-Herrero and Wilson, 2017), although urban areas are usually deemed the privileged location for KIBS, recent evidence shows that the increasing demand for knowledge-intensive services is driven by firms in spatially clustered industries in non-urban areas. In this regard, Lombardi et al. (2022) demonstrate that KIBS have a positive effect on manufacturing performance in both urban and non- urban areas, with the magnitude varying depending on contextual factors, and highlight the role of education level in determining this positive effect. By and large, such figures give a precise idea of the extensive reconfiguration of production organisation and business fields' borders experienced by industrialised countries, where the current wave of digital technologies, often referred to as Industry 4.0 or "the Fourth Industrial Revolution" (Schwab 2016), has indeed opened renewed opportunities as well as threats for manufacturing areas and SMEs.

Table 5.4 - Employment in KIBS by LMA type. Year 2015 (percentage values) (a) (b)

Type of LMA	LQ of KIBS	KIBS Employees on total employment (% values)	KIBS value added distribution (%values)	KIBS value added on Services value added (%values)
Urban LMAs	-	22.7	66.1	40.6
Local urban systems with high specialisation	1.5	28.6	41.2	47.5
Multi-specialised local urban systems	1.1	19.5	13.7	35.7
Predominantly local urban port systems	1.0	17.7	8.5	30.3
Non-specialised local urban systems	0.9	17.2	2.7	27.6
Other non- manufacturing systems	-	11.8	2.1	17.3
Tourist local systems	0.6	11.8	1.6	17.2
Agricultural vocation	0.6	11.7	0.5	17.7
Manufacturing - Made in Italy systems	-	14.5	16.2	29.2
Textile and clothing industries	0.8	14.8	3.7	30.1
Hides and leather industries	0.7	12.8	1.6	26.1
Machine manufacturing industries	0.8	14.7	4.0	31.6
Wood and furniture industries	0.8	15.3	3.1	30.4
Agro-food industries	0.8	14.3	2.8	26.0
Jewels, glasses and musical instruments	0.8	15.2	1.0	29.4
Heavy manufacturing systems	-	17.2	14.8	35.3
Local systems of transport industries	1.1	21.0	6.3	44.1
Metals production and processing	0.8	15.0	3.2	31.3
Construction materials industries	0.7	12.7	0.7	23.6
Petrochemical and pharmaceutical industries	0.9	16.9	4.6	31.8
Non-specialised local systems	0.6	11.4	0.9	18.4
ITALY	**1.0**	**18.6**	**100.0**	**36.1**

Source: Authors' processing of the Istat T-SBS Register data
(a) Service sectors include the following NACE Rev. 2 sections: "H", "I", "J", "L", "M", "N", "P", "Q", "R", "S",
(b) NACE codes for KIBS follow the territorial servitisation literature, which applies the Wong and He (2005) classification to identify economic activities that could be classified as services and are relevant to manufacturing companies. They establish the North American Industry Classification System (NAICS) codes relevant to servitisation theory. These codes are linked to the following economic activities: (1) IT and related services; (2) engineering and technical services; and (3) business and management consulting. See Lombardi et al. (2022) for the complete list, transcoded into NACE Rev.2 codes.

6. Concluding remarks

This paper has shown that, before the implementation of the T-SBS Register, estimates of value added at the local unit level were produced using a top-down method, in which the enterprise's value added was distributed pro quota to estimate the value added of each local unit. For enterprises responding to the survey on the economic accounts of large enterprises, the distribution was based on the functional regional unit distribution of labour costs; for those not responding, it was based on the distribution per local unit of the number of employees, taken from the BR-LU. By contrast, the methodology proposed here introduces innovations that greatly improve compliance with the requirements of the European Regulation, enhance the consistency of micro- and macro-level information on business structures, enrich the territorial detail of information available on businesses and local units, and increase analytical and interpretative potential. The exercise conducted to calculate and analyse value added and productivity at the municipal level - beyond the results obtained, which are of great interest and provide interesting research prospects - is an in vivo example of the potential offered by the new method of "punctual" (i.e., micro) estimation of value added at the finest territorial level.

Implications for research and practice are wide-ranging. The T-SBS register has already contributed to monitoring the effects of the COVID-19 pandemic, as territorial-level economic indicators now provide the most accurate data for analysing these effects (Istat 2020). Most importantly, the tool for interpreting reaction trajectories at the subnational level (Istat, 2021) and the emerging role of cities in the post-COVID era are notable (Armenise and Viesti 2021). Hence, business statistics at the territorial level can be of great use in policy-making, where workplace geography has attracted renewed attention (Faramondi et al. 2022; Franconi et al. 2017).

Finally, some directions for future research are needed. The T-SBS Register was deliberately conceived to improve the provision of data on business indicators by official statistics, with greater territorial granularity, thereby marking a difference from previous territorial microdata. It is a powerful tool for studying economic change through territorial-level analysis. In fact, this provision has improved academic research on the Italian landscape, identifying trajectories and sectors that are intertwined with local economies at different levels. Future research should strengthen collaboration with the national government and local institutions to develop business indicators to inform the design of targeted industrial policies that, more generally, address issues of local development.

References

Armenise, M., e G. Viesti. 2021. “Il sistema urbano come motore di rilancio del paese”. In Bellandi, M., I. Mariotti, e R. Nisticò (a cura di). *Città nel Covid. Centri urbani, periferie e territori alle prese con la pandemia*. Roma, Italia: Donzelli Editore.

Arnaldi, S., P.F. Aureli, A.M.M. Carucci, S. Curatolo, T. Di Francescantonio, A. Faramondi, R. Filippello, R. Nardecchia, A. Puggioni, A. Regano, R. Rizzi, G. Sacco, R. Sanzo, e P. Sassaroli. 2020. “Il trattamento delle fonti amministrative per la costruzione del sistema informativo sui risultati economici delle imprese italiane (Frame SBS)”. *Istat working papers,* N. 6/2020. Roma, Italia: Istat. https://www.istat.it/produzione-editoriale/il-trattamento-delle-fonti-amministrative-per-la-costruzione-del-sistema-informativo-sui-risultati-economici-delle-imprese-italiane-frame-sbs/.

Bigerna, S., C.A. Bollino, e P. Polinori. 2016. “Le determinanti della crescita e della disuguaglianza in Umbria prima della crisi”. In Bollino, C.A., e G. Espa. *Analisi e modelli di efficienza e produttività a livello territoriale*: 145-184. Milano, Italia: Franco Angeli.

Bollino, C.A., e G. Espa (a cura di). 2016. *Analisi e modelli di efficienza e produttività a livello territoriale*. Milano, Italia: Franco Angeli.

Chen, J., and J. Shao. 2000. “Nearest Neighbor Imputation for Survey Data”. *Journal of Official Statistics*, Volume 16, N. 2: 113-131. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/nearest-neighbor-imputation-for-survey-data.pdf.

Eurostat. 2013. *Manual on regional accounts methods*. Manuals and Guidelines. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/documents/3859598/5937641/KS-GQ-13-001-EN.PDF/7114fba9-1a3f-43df-b028-e97232b6bac5.

Faramondi, A. 2008. “Valore aggiunto comunale: integrazione tra fonti e approccio bottom-up”. *Rivista Internazionale di Scienze Sociali*, Volume 116, N. 2: 179-209.

Faramondi, A., V. De Giorgi, D. De Francesco, R. Di Manno, S. Lombardi, R. Nardecchia, R. Sanzo, V. Tomeo, e E. Trinca. 2018. "La stima del valore aggiunto a livello territoriale: il nuovo registro statistico 'frame SBS territoriale'". Paper presented at the *AISRe*, *XXXIX Conferenza Scientifica Annuale*, Bolzano/*Bozen*, 17-19 settembre 2018.

Faramondi, A., S. Lombardi, I. Straccamore, and D. De Francesco. 2022. "Report on the methodology for producing economic indicators at the local unit level and related data". Deliverable 4.3, Grant Agreement N. 882021, *Data collection for city and subnational statistics – Italy - 2019-IT-Subnational* (31st March 2022).

Franconi, L., D. Ichim, and M. D'Aló. 2017. "Labour Market Areas for Territorial Policies: Tools for a European Approach". *Statistical Journal of the IAOS*, Volume 33, N. 3: 585-591. https://doi.org/10.3233/SJI-160343.

Gigante, S., S. Pacini, F. Rossetti, e V. Talucci. 2019. "La qualità del lavoro dipendente nel settore privato. Una lettura integrata attraverso il registro RACLI". *Statistica e società,* Post (16 novembre 2019). https://www.rivista.sis-statistica.org/cms/?p=790.

Iuzzolino, G. 2004. "Costruzione di un algoritmo di identificazione delle agglomerazioni territoriali di attività manifatturiere". In Banca d'Italia. *Economie locali, modelli di agglomerazione e apertura internazionale*. *Nuove ricerche della Banca d'Italia sullo sviluppo sostenibile*. Atti del Convegno svoltosi a Bologna il 20 novembre 2003. Roma, Italia: Banca d'Italia.

Istituto Nazionale di Statistica - Istat. 2021. "Gli effetti territoriali della crisi economica". In Istat. *Rapporto sulla competitività dei settori produttivi. Edizione 2021*: Capitolo 4: 103-118. Letture Statistiche – Temi. Roma, Italia: Istat. https://www.istat.it/produzione-editoriale/rapporto-sulla-competitivita-dei-settori-produttivi-edizione-2021-2/.

Istituto Nazionale di Statistica - Istat. 2020. *Dati comunali su Imprese, addetti e risultati economici delle imprese incluse in settori "attivi" e "sospesi" secondo i decreti governativi approvati a marzo 2020 (DPCM dell'11/03/2020 e DM Mise 25/03/2020).* Roma, Italia: Istat. https://www.istat.it/notizia/dati-comunali-su-imprese-addetti-e-risultati-economici-delle-imprese-incluse-in-settori-attivi-e-sospesi-secondo-i-decreti-governativi-approvati-a-marzo-per-l/.

Istituto Nazionale di Statistica - Istat. 2015a. *La nuova geografia dei sistemi locali*. Letture Statistiche - Territorio. Roma, Italia: Istat. https://www.istat.it/it/files/2015/10/La-nuova-geografia-dei-sistemi-locali.pdf.

Istituto Nazionale di Statistica - Istat 2015b. *I distretti industriali. Anno 2011*. Statistiche Report. Roma, Italia: Istat. https://www.istat.it/comunicato-stampa/i-distretti-industriali-anno-2011/.

Istituto Nazionale di Statistica - Istat 2015c. *Classificazioni dei sistemi locali*. Statistiche Sperimentali. Roma, Italia: Istat. https://www.istat.it/statistica-sperimentale/movimento-alberghiero-3/.

Istituto Nazionale di Statistica - Istat. 2005. *8° Censimento generale dell'industria e dei servizi. Distretti industriali e sistemi locali del lavoro 2001*. Roma, Italia: Istat. https://www.istat.it/it/files/2011/01/Volume_Distretti1.pdf.

Kalton, G., and D. Kasprzyk. 1982. "Imputing for missing survey responses". In American Statistical Association - ASA. *Proceedings of the section on Survey Research Methods*: 22-31. Alexandria, VA, U.S.: ASA. http://www.asasrms.org/Proceedings/y1982f.html.

Lafuente, E., Y. Vaillant, and F. Vendrell-Herrero. 2017. "Territorial servitization: Exploring the virtuous circle connecting knowledge-intensive services and new manufacturing businesses". *International Journal of Production Economics*, Volume 192: 19-28. https://doi.org/10.1016/j.ijpe.2016.12.006.

Little, R.J.A. 1988. "Missing-Data Adjustments in Large Surveys". *Journal of Business & Economic Statistics*. Volume 6, N. 3: 287-296. https://doi.org/10.2307/1391878.

Lombardi S., E. Santini, and C. Vecciolini. 2022. "Drivers of territorial servitization: an empirical analysis of manufacturing productivity in local value chains". *International Journal of Production Economics*, Volume 253. https://doi.org/10.1016/j.ijpe.2022.108607.

Lombardi, S., and F. Sforzi. 2016. "Chinese manufacturing entrepreneurship capital: evidence from Italian industrial districts". *European Planning Studies*, Volume 24, N. 6: 1118-1132. https://doi.org/10.1080/09654313.2016.1155538.

Luzi, O., M. Di Zio, U. Guarnera, A. Manzari, T. De Waal, J. Pannekoek, J. Hoogland, C. Templeman, B. Hulliger, and D. Kilchman. 2007. "Recommended practices for editing and imputation in cross-sectional business surveys". *European Project EDIMBUS*. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/documents/64157/4374310/30-Recommended+Practices-for-editing-and-imputation-in-cross-sectional-business-surveys-2008.pdf.

Luzi, O., and R. Monducci. 2016. "The new statistical register "Frame SBS": overview and perspectives". *Rivista di Statistica Ufficiale/Review of official statistics*, N. 1/2016: 5-14. Roma, Italy: Istat. https://www.istat.it/en/publication/rivista-di-statistica-ufficiale-review-of-official-statistics-12016/.

Menghinello, S., A. Faramondi, and T. Laureti. 2020. "The future role of official statistics in the business data arena". *Statistical Journal of the IAOS*, Volume 36, N. 2: 519-533. https://doi.org/10.3233/SJI-190557.

Nordbotten, S. 2010. "The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries". In Carlson, M., H. Nyquist, and M. Villani (Eds.). *Official statistics: methodology and applications in honour of Daniel Thorburn*: Chapter 17: 205-223. https://officialstatistics.wordpress.com/contents/.

Organisation for Economic Co-operation and Development - OECD, and European Commission. 2020. *Cities in the World: A New Perspective on Urbanisation*. OECD Urban Studies. Paris, France: OECD Publishing. https://doi.org/10.1787/d0efcbda-en.

Rubin, D.B. 1986. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations". *Journal of Business & Economic Statistics*, Volume 4, N. 1: 87-94. https://doi.org/10.2307/1391390.

Schwab, K. 2016. *The Fourth Industrial Revolution*. Geneva, Switzerland: The World Economic Forum.

Sforzi, F. 2009. "The Empirical Evidence of Industrial Districts in Italy". In Becattini, G., M. Bellandi, and L. De Propris (Eds.). *A Handbook of Industrial Districts*: Part III, Chapter 25. Cheltenham, UK: Edward Elgar Publishing.

Vendrell-Herrero, F., and J.R. Wilson. 2017. "Servitization for territorial competitiveness: taxonomy and research agenda". Competitiveness Review, Volume 27, N. 1: 2-11. https://doi.org/10.1108/CR-02-2016-0005.

Detecting under-reporting of value added in National Accounts

Federico Sallusti[1], Luciano Cavalli[1], Agata Maria Madia Carucci[1]

Abstract

The non-observed economy accounted for 9.5% of GDP and 10.5% of value added generated by the Italian business system in 2020. Under-reporting of value added is the main component, amounting to about 80 billion euros. This paper presents the ROC-Is method, developed by the Italian National Institute of Statistics (Istat), for measuring under-reporting at the micro-level. The method is based on ROC analysis and allows the profiling of tax-evading firms using a large set of economic and structural indicators. The method overcomes the limitations of the former "Franz" approach. Indeed, ROC-Is maintains its conceptual suitability regardless of firm size, eliminates any anti-cyclical bias, and uses a larger set of structural and economic information.

Keywords: National Accounts, underground economy, non-observed economy, ROC analysis.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2024.02.

1 Federico Sallusti (fsallusti@istat.it); Luciano Cavalli (lcavalli@istat.it); Agata Maria Madia Carucci (carucci@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

The Non-Observed Economy (NOE, hereinafter) comprises economic activities that escape direct statistical observation. Its impact on the economy varies widely across countries (OECD 2002; UNECE 2021; Fernandes 2022). However, it represents a significant share of the economy in many countries, including Italy, where in 2020 the NOE amounted to about 174 billion euros, accounting for 9.5% of GDP and 10.5% of value added generated by the business system (Istat 2024).

Estimates of NOE are necessary to ensure the exhaustiveness of National Accounts (NA) figures and to enable reliable comparisons of Gross Domestic Product (GDP) and Gross National Income (GNI) over time and across countries (Gyomai and van de Ven 2014). This is particularly relevant for European Union (EU) member states, as European policies and control mechanisms (such as the Excessive Deficit Procedure) are largely based on these indicators (European Parliament and Council of the European Union 2019).

Stressing the importance of measuring NOE, Eurostat (2021) recommended, from 2016, that EU member states use the so-called Tabular Approach for Exhaustiveness (TAE) as the framework for presenting separate measures of NOE components. Accordingly, National Statistics Offices (NSOs) of EU countries had to develop or improve their methods to obtain segmented estimates of NOE.

Underground production is the largest component of NOE in all EU countries; in Italy, it accounted for 90% of NOE in 2020 (157 billion euros). It is, in turn, mainly composed of under-reporting of value added, which is connected with false declarations aimed at under-reporting production and/or over-reporting costs to reduce tax payments on profits (slightly less than 80 billion euros in Italy), and of the value added generated by the employment of an unregistered workforce (about 62 billion euros in Italy)[2].

This work presents the ROC-Indicators (ROC-Is) method, recently developed by the Italian National Institute of Statistics (Istat) to measure under-reporting of value added at the micro level. The method is based on ROC analysis, which is widely used in medicine (Kumar and Indrayan 2011), machine learning (Majnik

2 The remaining part of underground production is composed of tips, unregistered rents, VAT fraud and balancing procedure between supply and demand. For further details, see the methodological note in Istat (2024).

and Bosnic 2013) and the natural sciences (Warnock and Peck 2010), and has recently become a relevant tool in economics (Costa et al. 2019, 2022a, 2022b; Sallusti 2025). In particular, ROC analysis is used to determine the threshold that classifies firms as under-reporting (or not) based on the value of a composite indicator summarising the main economic and structural characteristics of homogeneous firms by economic activity, size and geographical location. The same threshold is then used to adjust the value added of under-reporting firms, thereby obtaining a measure of the hidden value added.

The ROC-Is method uses the extensive dataset provided by the Frame-SBS archive (Luzi and Monducci 2016), which includes comprehensive economic and structural information for the entire population of Italian firms (about 4.3 million units in 2020). It also allows consideration of a wide range of characteristics when defining the profiles of business units in light of their fiscal behaviour.

The ROC-Is method was developed to address the main shortcomings of the preceding approach to measuring under-reporting of value added. Indeed, according to the “Franz” method (Franz 1983), under-reporting of value added was measured by comparing the declared profits of firms (entrepreneurs) with an opportunity cost proxied by the wage the entrepreneur would be able to earn working as an employee in a similar production context (in terms of skills, industry and firm typology). In particular, under-reporting of value added is measured by imposing the constraint that profits cannot be lower than the opportunity cost for each firm.

The “Franz” method suffers from three key limitations. First, it becomes conceptually unsuitable as enterprise size increases. Indeed, the behavioural assumption underlying the approach is reliably applicable only to very small firms (i.e., self-employed). Second, it is, by construction, anti-cyclical, as wages tend to be less sensitive than profits to the business cycle. Third, information on firms’ structural and economic characteristics is taken into account only to a limited extent.

The ROC-Is method addresses all the underlying shortcomings of the old-fashioned “Franz” approach. First, it remains conceptually suitable regardless of firm size, as it does not assume any specific “individual behaviour” among entrepreneurs. Second, it compares firms’ relative performance, implicitly eliminating the “forced” anti-cyclical behaviour of the “Franz” method.

Third, it uses a large set of structural and economic information to build profiles of business units.

The remaining work is organised as follows. The second Section presents the ROC-Is procedure and the database used in the analyses. The third Section presents the results obtained over the past five years. The fourth Section concludes.

2. ROC-Indicator's procedure

The magnitude and characteristics of non-observed events are difficult to measure using direct methods. Generally, they are approached using indirect estimates based on observable data, with the aim of identifying "abnormal" behaviours and possibly attributing this deviation from "normality" to the non-observed factor.

In this context, the reliability and suitability of indirect approaches depend on the validity of the conceptual framework used to interpret non-observed events, and on the characteristics of the data used to identify and measure them. The former is crucial for determining whether, and to what extent, deviations from "normality" are attributable to non-observed behaviour, while the latter is relevant for ensuring coherence between the conceptual model and its information counterpart.

The ROC-Is procedure is an indirect method for detecting and adjusting for the under-reporting of value added by firms. The conceptual framework is based on the idea that fiscal misbehaviour can be detected by analysing inconsistencies in firms' economic and structural data. In this context, ROC analysis allows the determination of the threshold at which such inconsistencies can be probabilistically attributed to under-reporting.

The rest of the Section is organised as follows. The first paragraph is devoted to stress manipulation and data processing. The second provides some insight into ROC analysis. The third outlines the different stages of the ROC-Is method.

2.1 Data

The possibility of defining a method for measuring under-reporting at the micro level implies the availability of a large dataset on the structural and economic characteristics of business units. In recent years, Istat developed Frame-SBS (Luzi and Monducci 2016), an archive that integrates survey data and administrative records and includes comprehensive economic and structural information for the entire population of Italian firms (about 4.3 million in 2020).

The ROC-Is method is applied to the subset of "economically relevant" firms with fewer than 100 workers that are not subject to specific non-treatability conditions. In 2020, about 2.5 million firms were considered, generating 310 billion euros of (regular) value added and employing 8.9 million workers. In this context, Table 2.1 presents descriptive statistics for the dataset.

Table 2.1 - Descriptive statistics of the dataset by economic activity, legal form and size class. Year 2020

	Firms		Value added		Workers	
	Units	%	Billion euros	%	Thousands	%
Economic activity						
Manufacturing, mining and quarrying and other industry	274,052	10.9	82.9	26.7	1,919.2	21.6
Construction	330,582	13.1	37.2	12.0	1,031.1	11.6
Wholesale and retail trade, transportation and storage, accommodation and food service activities	987,248	39.2	98.2	31.7	3,610.8	40.7
Other market services	647,619	25.7	68.9	22.2	1,547.3	17.4
Education and human health	159,697	6.3	15.9	5.1	385.7	4.3
Other services	119,809	4.8	6.9	2.2	381.0	4.3
Legal forms						
Sole trader	1,188,296	472	82.9	26.7	1,919.2	21.6
Partnerships	491,599	19.5	37.2	12.0	1,031.1	11.6
Joint-stock company	10,516	0.4	98.2	31.7	3,610.8	40.7
Limited liability company	797,095	31.6	68.9	22.2	1,547.3	17.4
Cooperatives	30,543	1.2	15.9	5.1	385.7	4.3
Other	958	0.0	6.9	2.2	381.0	4.3
Size class						
0 - 1	1,038,584	41.2	45.6	14.7	925.7	10.4
2 - 5	1,112,777	44.2	73.7	23.8	3,150.4	35.5
6 - 9	197,847	7.9	42.9	13.9	1,405.4	15.8
10 - 19	118,590	4.7	60.0	19.3	1,547.6	17.4
20 - 49	42,169	1.7	57.1	18.4	1,236.6	13.9
50 and over	9,040	0.4	30.7	9.9	609.3	6.9
Total	**2,519,007**	**100.0**	**309.9**	**100.0**	**8,875.1**	**100.0**

Source: Authors' processing of Istat data

Based on the extensive information contained in the Frame-SBS, a large set of indicators has been considered to describe firms' economic behaviour. In particular, for each firm, indicators have been defined to establish a (supposed) monotonic relationship with the likelihood of under-reporting.

Indicators relate to three main areas of the economic behaviour of business units.

Performance and profitability:

- Value added-per-worker;
- Earnings Before Interest, Taxes, Depreciation, and Amortisation (EBITDA)-per self-employed;
- EBITDA-turnover ratio;
- Profit-turnover ratio;
- ROI;
- Profit-EBITDA ratio.

Structure of costs:

- Labour cost-total cost ratio;
- Management cost-total cost ratio;
- Goods/services cost-total cost ratio;
- Structural cost-total cost ratio;
- Goods/services costs on inventory rotation.

Structure of employment:

- Temporary workers-total workers ratio;
- Outworkers-total workers ratio;
- Self-employed workers-total workers ratio.

To maintain homogeneity in firms' economic and structural characteristics, strata are defined by the following dimensions: (1) industry; (2) size class; and (3) geographical areas[3].

2.2 ROC analysis

ROC analysis falls within the category of classification models. In particular, it allows us to define a cut-off value along a target indicator that should permit us to classify an observation with respect to a given binary

3 In particular, the procedure is applied to strata defined by four macro-sectors (manufacturing, construction, trade, transport and accommodation and restaurants, other services), five classes (1-2 workers, 2-5 workers, 5-10 workers, 10-20 workers, 20-100 workers), and two geographical areas (Nord-Centre, and South and Islands).

characteristic. Following Fawcett (2006), classification models can give the following four possible outcomes (which are also shown in the "confusion matrix" in Figure 2.1)[4]:

- True Positives (TP, a): positive observations are correctly classified as positive by the model;
- False Negatives (FN, b): positive observations are erroneously classified as negative by the model;
- False Positives (FP, c): negative observations are erroneously classified as positive by the model;

True Negatives (TN, d): negative observations are correctly classified as negative by the model.

Figure 2.1 - Confusion matrix

		Estimated classification	
		1	0
True classification	1	**TP** *(a)*	**FN** *(b)*
	0	**FP** *(c)*	**TN** *(d)*

Source: Authors' processing

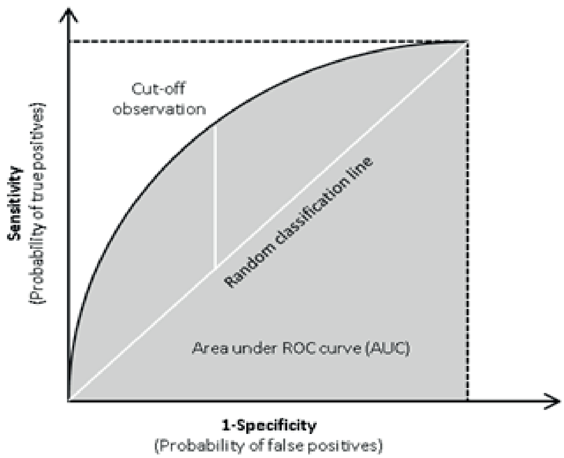
The validity of a classifier can be measured using two main metrics. Sensitivity, which represents the probability of detecting true-positive cases (in terms of Figure 2.1: *a/(a+b))*. Specificity, which is the probability of detecting true-negative cases (in terms of Figure 1: *d/(c+d))*. The latter is usually considered in its reciprocal expression (1-Specificity), which measures the probability of false positives.

4 For an in-depth description of the ROC analysis, see Costa et al. (2019, 2022a, 2022b).

Once a classifier is applied, the ROC curve in Figure 2.2 plots each observation in the space of Sensitivity and 1-Specificity, illustrating the trade-off between the probabilities of detecting true positives and false positives across all possible cut-off points (Kumar and Indrayan 2011).

In this context, the area under the ROC curve (AUC, the grey area in Figure 2.2) quantifies how much more efficient the clustering produced by the given model is than a purely random classification (the 45° line).

Figure 2.2 - ROC curve



Source: Authors' processing

Considering Figure 2.2, the threshold (cut-off point) that classifies the observation, based on the shape of the given indicator, can be set using the following condition:

$$\widehat{Cut} = h * Sensitivity - (1-h) * (1 - Specificity) \qquad (1)$$

where h and (1 - h) represent the relative weights assigned to the different elements of the trade-off in clustering. In particular, when $h = 0.50$, a “neutral” identification is obtained – the so-called Youden (1950) index[5]. If h is set to a value greater than 0.5, true positives are prioritised over false positives. If h is set to a value lower than 0.5, avoiding false positives is considered more important than identifying true positives.

5 The Youden (1950) index permits maximisation of the vertical distance between the ROC curve and the random classification line.

Thus, the final clustering obtained using ROC analysis depends on two main elements: the first concerns the indicator's ability to explain the distribution of the binary variable, which determines the shape of the ROC curve; the second concerns the relative weights assigned to classification errors.

2.3 The ROC-Is method

The ROC-Is method comprises three stages. The first involves constructing the composite indicator that summarises firms' profiles and defining the binary variable. The second uses ROC analysis to determine the threshold that classifies firms as under-reporting or not (i.e., identification). The third determines the amount of adjustment needed to make the value added of the under-reporting firm consistent with the condition of no under-reporting (i.e., the adjustment).

2.3.1 The composite indicator and the suspicion of under-reporting

In the ROC-Is method, the binary variable is represented by the suspect of under-reporting, while the classifier is represented by a composite indicator that summarises the information provided by the structural and economic indicators defined in paragraph 2.1[6].

The binary variable representing the proxy for under-reporting (indicating a "suspect" of tax evasion) is defined for the *i-th* firm in the *s-th* stratum by comparing the suitable performance indicator with the stratum average. Accordingly, firms are classified as "suspect" (or not) based on whether their performance indicator is below (or above) the stratum average.

The composite indicator summarising the profile of firms is constructed in three steps. In the first step, for the *s-th* stratum, a logit model is estimated with the proxy for "suspect" as the dependent variable and the full set of indicators as covariates. The model's results allow identification of the five most informative indicators. In the second step, a factor analysis is performed to obtain three factors that summarise the information from the relevant

6 The need to define a composite indicator that summarises the different characteristics of business units is connected with the features of the ROC analysis, which can be performed (to determine the threshold) only using a single variable.

indicators.Finally, in the third step, the factors are grouped to define, for the *i-th* firm in the *s-th* stratum, the composite indicator Z_i, which takes the following form (the subscript for the stratum in the notation is omitted for the sake of simplicity):

$$Z_i = \sum_j \omega_j F_{i,j} \qquad (2)$$

where ω_j are the shares of explained variance for each factor, and the *j-th* factor is:

$$F_{i,j} = \sum_k \gamma_{k,j} \alpha_{i,k,j} \qquad (3)$$

where $\alpha_{i,k,j}$ is the *k-th* indicator for the *i-th* in the *j-th* factor, and $\gamma_{k,j}$ is the relative loading.

2.3.2 Identification

The identification of under-reporting firms is based on ROC analysis. In particular, starting from a logit model, ROC analysis identifies a threshold value across the covariate distribution, which allows classification of observations with respect to the binary response variable, taking into account the relative weight of possible misclassifications (false positives or false negatives).

In the context of this work, ROC analysis is used to discriminate between firms that under-report and those that do not, based on the relationship between the composite indicator representing the economic behaviour of business units (Z_i) and the proxy representing the "suspect" of under-reporting.

When clustering firms based solely on the results of the proxy-vs-composite indicator logit model, applying ROC analysis offers two main advantages. The first is a statistical rather than a subjective definition of the clustering threshold for positive and negative observations[7]. The second is linked to the implicit correction that ROC analysis applies to the proxy variable's informative capability (Sallusti 2025).

7 The logit model provides, for each observation, a probability of belonging to the groups defined by the binary variable. In this context, clustering of observations should depend on the probability threshold the researcher sets. ROC analysis provides a method for statistically defining this cut-off value by optimising it to take into account the different weights the researcher is willing to assign to the different types of error (false negatives and false positives) (Fawcett 2006).

The identification procedure is composed of three steps.

In the first step, the following logit model is estimated (the subscript for the stratum is omitted for simplicity):

$$Prob(proxy = 1) = \Lambda(\beta Z_i) \qquad (4)$$

where Λ is the cumulative distribution of the logistic function, β is the estimated parameter and Z_i is the composite indicator.

The second step is devoted to defining the threshold based on which firms are classified as under-reporting or not. In particular, along the ROC curve generated by the logistic model in Equation (4) (Figure 2.2), the cut-off observation (which represents the threshold for the given stratum) is defined by Equation (1) with h set to 0.5[8].

In the third step, firms in a given stratum are finally classified as under-reporting or not, based on a comparison between the value of their composite indicator Z_i and the value of the composite indicator for the threshold observation. In particular, assuming $\bar{Z}$ as the value of the composite indicator for the cut-off observation, other firms are classified as under-reporting if $Z_i < \bar{Z}$.

2.3.3 Adjustment

The value-added adjustment for under-reporting firms is derived using the identification information. In particular, for the *i-th* under-reporting firm, the adjustment is obtained by increasing the composite indicator (leveraging the value-added-per-worker indicator) to the threshold value defined by the ROC analysis. This way, each under-reporting firm is brought to the threshold value, so it is no longer considered under-reporting.

For the *i-th* under-reporting unit in the *s-th* stratum, the following condition holds (the subscript for the stratum is omitted in the notation for simplicity):

$$\sum_j \omega_j F_{i,j} < \bar{Z} \qquad (5)$$

8 Assuming the "neutral" selection seems to be the more suitable choice in this context, since there is no strong a priori evidence about the "right" evidence with which to classify business units as under-reporting or not. The choice of different values of h might be useful if, for example, the method were used for audit purposes (where auditors might be interested in increasing the efficiency of controls by reducing h, to submit to audit only firms with stronger evidence of tax evasion).

where the first component is the value of the composite indicator for the *i-th* under-reporting firm, and $\bar{Z}$ is the threshold value for classification in the *s-th* stratum.

The adjustment is thus obtained using the following condition:

$$\tilde{\alpha}_{h,i}: \sum_j \omega_j F_{i,j} = \bar{Z} \qquad (6)$$

In particular, the composite indicator is incremented by leveraging on the value added-per-worker indicator ($\tilde{\alpha}_{h,i}$) as show in the following equation:

$$\tilde{\alpha}_{h,i} = \frac{\bar{Z} - \sum_j \omega_j \gamma_{j,-h} \alpha_{-h,i}}{\sum_j \omega_j \gamma_{j,h}} \qquad (7)$$

where, for the given stratum, $\tilde{\alpha}_{h,i}$ is the adjusted value-added-per-worker indicator for the *i-th* under-reporting firm, $\bar{Z}$ is the threshold, ω_j are the weights used to aggregate factors into the composite, $\gamma_{j,n}$ are the loadings representing the weight of each indicator α_k in the definition of the *j-th* factor (the subscript *h* indicates the value-added-per-worker indicator, while the subscript -*h* indicates other indicators).

Finally, the level of adjustment y_i can be obtained, for the *i-th* under-reporting firm in the *s-th* stratum, as:

$$y_i = (\tilde{\alpha}_{h,i} - \alpha_{h,i}) * N_i \qquad (8)$$

where $\tilde{\alpha}_{h,i}$ is the adjusted value-added-per-worker indicator, $\alpha_{h,i}$ is the declared value-added-per-worker and N_i is the number of workers of the *i-th* firm.

Equation (7) implies that the magnitude of the adjustment (i.e., the amount of under-reporting) depends on three elements. The first is the level of the threshold $\bar{Z}$, which represents the contextual conditions in which the under-reporting firm operates (in terms of industry, size class, and geographical areas). Indeed, the difference between $\bar{Z}$ and the value of the composite indicator for the *i-th* firms (Z_i) can be interpreted as a proxy for the deviation of the under-reporting unit from "normality", i.e., the minimum requirements to be included in the non-under-reporting class in its stratum. In this context, the greater the distance, the larger the adjustment.

The second element is represented by the rest of the numerator ($\sum_j \omega_j \gamma_{j,-h} \alpha_{-h,i}$), which captures the effect of the other indicators (than value added per worker) (α_{-h}) on the distance between Z_i and the threshold. The greater their influence, the lower, *ceteris paribus*, the amount of the adjustment.

The third element is the denominator ($\sum_j \omega_j \gamma_{j,h}$), which represents the influence of the value added per worker (α_h). In this case, the higher the response, the smaller the adjustment.

The adjustment for *i-th* under-reporting in the business unit will therefore depend on both meso-economic and individual characteristics, with the former summarised by the distance between the threshold and the firm, and the latter by the relative relevance of value added per worker and the other variables in the composite indicator. This confirms that the ROC-Is method permits measurement of under-reporting by taking into account not only sectoral and general meso- and macro-level elements, but also the individual economic structures of firms.

3. Results

In this Section, the main results of the measurement of under-reporting are presented for the 2016-2020 period. Results are shown by sector, legal form, and size class.

Table 3.1 reports the number, employment, and value added of business units submitted to the ROC-Is method, as well as their incidence relative to the entire Italian business system. From 2016 to 2020, the number of business units decreased from 2,8 to 2,5 million (from 63.3% to 57.1% of the total population of firms). Employment and value added were more stable over the years: the incidence in terms of employment ranged between 55.0% and 53.0%, while the incidence in terms of value added ranged between 51.2% and 48.6%. The reduction in the number of firms in the last two years is related to very small firms and is due to the new definition of enterprises under "flat-tax rate" regimes (see footnote 6) introduced in 2019.

Table 3.1 - Business units submitted to the ROC-Is method by number of firms, workers, and value added. Years 2016-2020 (absolute values, percentages and billions of euros)

	2016	2017	2018	2019	2020
Number of firms	2,819,073	2,806,006	2,838,944	2,597,083	2,519,007
Incidence on total economy	*63.3*	*63.0*	*63.6*	*59.5*	*57.1*
Number of workers	8,943,422	9,087,953	9,315,554	9,151,456	8,875,100
Incidence on total economy	*55.0*	*54.6*	*55.1*	*53.8*	*53.0*
Value added (billion)	388.59	397.27	418.68	417.25	365.36
Incidence on total economy	*51.2*	*50.7*	*51.7*	*50.2*	*48.6*

Source: Authors' processing of Istat data

Table 3.2 shows the incidence of under-reporting, measured as the number of units relative to the total population of firms subject to the method. Most firms are identified as under-reporting: the overall incidence ranges from 55.8% in 2016 to 52.5% in 2020. Results are relatively stable during 2016-2019 (from 54.2% to 55.8%), while in the last year the incidence decreases by approximately 3 percentage points (to 52.5%) due to the effects of the COVID-19 pandemic.

Table 3.2 - Under-reporting units by economic activity, legal form, and size class. Years 2016-2020 (percentage values) (a)

	2016	2017	2018	2019	2020
Economic activity					
Manufacturing, mining and quarrying and other industry	56.2	55.1	55.8	55.1	53.4
Construction	60.0	57.3	52.8	53.5	48.5
Wholesale and retail trade, transportation and storage, accommodation and food service activities	64.9	62.3	64.4	63.4	60.2
Other market services	43.4	42.7	45.0	44.3	41.6
Education and human health	29.7	28.4	27.6	31.7	36.0
Other services	72.0	75.6	72.8	80.4	78.3
Legal forms					
Sole trader	56.9	54.8	57.3	55.1	52.2
Partnerships	67.7	65.5	65.6	69.6	65.2
Joint-stock company	10.5	13.1	12.0	11.3	14.1
Limited liability company	44.0	44.4	42.4	46.3	45.3
Cooperatives	59.4	60.6	59.7	59.6	59.8
Other	22.6	30.8	26.9	27.3	21.5
Size class					
0 - 1	51.2	48.7	48.6	47.9	44.9
2 - 5	61.9	62.0	63.0	63.8	60.6
6 - 9	56.4	51.1	56.0	56.5	55.9
10 - 19	49.2	46.3	48.0	46.4	44.6
20 - 49	36.4	36.9	37.8	36.0	35.4
50 and over	35.0	35.3	35.1	31.5	32.1
Total	**55.8**	**54.2**	**54.9**	**55.3**	**52.5**

Source: Authors' processing of Istat data
(a) According to footnote 3, social cooperatives are excluded.

Considering the classification by economic activity, in 2020, under-reporting units are concentrated in Other services (78.3%) and Trade, transportation, hotels and restaurants (60.2%). Industry shows an incidence of 53.4%, while Education and human health (36%), Other market services (41.6%), and Construction (48.5%) are characterised by an incidence below 50%.

During 2016-2020, sectors show different trends. Industry, Trade, transportation, hotels and restaurants, and Other market services are almost stable in 2016-2019 (respectively 56.2% to 55.1%, 64.9% to 63.4%, and 43.4% to 44.3%), while decreasing in 2020 (1.7, 3.2, and 2.7 percentage points). Construction continued to decrease (by 11.5 points over the 2016-2020 period), while Education, human health, and Other services increased (by more than 6 points over 5 years).

Considering the breakdown by legal forms, Partnerships (65.2%) and Cooperatives (59.8%) show an incidence higher than the total average (52.5%), whereas Joint-stock companies are characterised by the lowest incidence (14.1%). Over the entire period, the weight remains nearly stable for Limited Liability Companies and Cooperatives. Sole traders and Partnerships decrease by 4.7 and 2.5 points, respectively, while Joint-stock companies increase by 3.6 points.

Finally, the breakdown by size class showed a more homogeneous pattern. Across all classes, the incidence declined over the 2016-2020 period, with the 0-1-worker (6.3 points) and 10-19-worker (4.6 points) classes showing the largest declines. The 2-5-worker and 6-9-worker classes had the highest incidence rates (60.6% and 55.9%, respectively, in 2020).

The ROC-Is method estimates, for each under-reporting firm, the amount of adjustment, which, over the period considered, ranged from 55.4 billion euros (in 2020) to 65.0 billion euros (in 2018) (about two-thirds of the total adjustment for under-reporting in Italy).

Table 3.3 shows the incidence of the adjustment, defined as the ratio of under-reporting to total value added (including the adjustment). Overall, the incidence of under-reporting declined over the period, from 16.7% in 2016 to 15.2% in 2020.

Across economic activities, all sectors except Other services (+0.7 percentage points) recorded a lower incidence of adjustment in 2020 than in 2016. In this context, Other market services reported the largest decline in the incidence of under-reporting (-2.8 percentage points between 2016 and 2020), while reductions in adjustment results were also notable in Industry (-1.3 percentage points) and Construction (-1.2 percentage points). Over the whole period, the highest incidence of under-reporting was observed in Other market services (between 34.0% and 36.2%) and in Trade, transportation, hotels and restaurants (between 19.4% and 20.9%), while Industry showed the lowest incidence (between 10.0% and 11.4%).

Breaking down the results by legal form, Cooperatives had the highest incidence in 2020 (25.3%), followed by Partnerships (22.5%) and Sole traders (19.8%), all above the Italian average (15.2%). The lowest incidence was recorded for Joint-stock companies (2.6%) and Limited liability companies (11.9%). Over the 5-year period, Cooperatives decreased their incidence by 2.9 percentage points, Sole traders by 1.7 points, and Partnerships by 1.5 points. By contrast, Joint-stock companies and Limited liability companies increased their incidence by 1.4 and 0.3 points, respectively.

By size class, the incidence of adjustment was highest in the 2-5-worker class (21.4%), followed by the 0-1-worker class (20%) and the 6-9-worker class (17.2%), indicating a clear inverse relationship between the prevalence of under-reporting and firm size.

Table 3.3 - Adjustment by economic activity, legal form and size class. Years 2016-2020 (percentage values) (a)

	2016	2017	2018	2019	2020
Economic activity					
Manufacturing, mining and quarrying and other industry	11.4	11.0	10.5	10.0	10.1
Construction	14.6	14.5	13.8	12.9	13.4
Wholesale and retail trade, transportation and storage, accommodation and food service activities	20.9	20.7	19.8	19.4	20.4
Other market services	14.6	13.8	13.3	12.4	11.8
Education and human health	11.7	10.5	10.0	10.9	11.1
Other services	35.3	36.2	34.0	34.1	36.0
Legal forms					
Sole trader	21.5	21.0	21.8	20.5	19.8
Partnerships	24.0	23.4	22.3	22.7	22.5
Joint-stock company	1.2	1.8	1.7	1.7	2.6
Limited liability company	11.6	11.8	10.3	10.9	11.9
Cooperatives	28.2	29.1	26.8	25.1	25.3
Other	22.2	24.0	12.0	19.6	13.9
Size class					
0 - 1	21.8	20.6	20.1	21.0	20.0
2 - 5	22.5	22.9	21.9	21.4	21.4
6 - 9	17.6	17.2	16.9	17.1	17.2
10 - 19	11.7	11.7	10.9	10.0	10.5
20 - 49	8.7	8.9	8.3	8.1	8.8
50 and over	8.8	8.5	7.5	7.0	7.6
Total	**16.7**	**16.3**	**15.5**	**15.1**	**15.2**

Source: Authors' processing of Istat data
(a) According to footnote 3, social cooperatives are excluded.

4. Concluding remarks

The aim of this work was to present the procedure developed by Istat for estimating the under-reporting of value added. The new ROC-Is method was designed to overcome the limitations of the old “Franz” approach.

In particular, the conceptual reliability of the ROC-Is method is independent of firms’ size; the results are not artificially anti-cyclical; and the procedure exploits a large share of the vast information contained in the Frame-SBS archive.

The ROC-Is method applies ROC analysis to a composite indicator that profiles firms’ economic behaviour and to a classification variable representing the “suspect” of under-reporting. This model allows both the identification of under-reporting firms and the adjustment of their value added.

Both identification and adjustment are conceptually linked to the ability to classify business units according to each unit’s position relative to a threshold, which, in turn, depends on the distribution of firms’ economic behaviour (as represented by the composite indicator). In this respect, the method permits the classification of business units by taking into account both their specific characteristics and the stratum's general trend and structure, thereby avoiding any subjective evaluation.

The ROC-IS method provides firm-level estimates of under-reporting, enabling granular analysis of the phenomenon. As shown in the paper, the method’s results can be applied from different perspectives based on business-unit characteristics. Sector, legal form, and size class are among the perspectives that can be used to examine the profile of business units that underreport value added. Indeed, granular measurement of under-reporting also allows analyses in light of relationships with firms’ and sectors’ performance in terms of productivity, profitability, and concentration. Furthermore, governance and other strategic and organisational characteristics could be used to profile the fiscal behaviour of business units.

The ROC-Is method, therefore, not only substantially improves the measurement of under-reporting in the Italian economy, with positive effects on the reliability and exhaustiveness of National Accounts figures, but also provides a reference point for the micro-analysis of fiscal misbehaviour within the Italian business system.

References

Breusch, T. 2016. "Estimating the Underground Economy using MIMIC models". *Journal of Tax Administration*. Volume 2, N. 1: 41-72. https://jota.website/jota/article/view/137.

Costa, S., F. Sallusti, C. Vicarelli, and D. Zurlo. 2022a. "Tech on the ROC: export threshold and technology adoption interacted". *Small Business Economics*, Volume 59, N. 4: 1593–1611. https://doi.org/10.1007/s11187-021-00581-7.

Costa, S., F. Sallusti, C. Vicarelli, and D. Zurlo. 2022b. "Firms' solidity before an exogenous shock: Covid-19 pandemic in Italy". *Economic Analysis and Policy*, Volume 76: 946-961. https://doi.org/10.1016/j.eap.2022.10.007.

Costa, S., F. Sallusti, C. Vicarelli, and D. Zurlo. 2019. "Over the ROC methodology: Productivity, economic size and firms' export thresholds". *Review of International Economics*, Volume 27, N. 3: 955-980. https://doi.org/10.1111/roie.12405.

European Parliament, and Council of the European Union. 2019. *Regulation (EU) 2019/516 of the European Parliament and of the Council of 19 March 2019 on the harmonisation of gross national income at market prices and repealing Council Directive 89/130/EEC, Euratom and Council Regulation (EC, Euratom) No 1287/2003 (GNI Regulation)*, Brussels, Belgium: European Parliament. https://eur-lex.europa.eu/eli/reg/2019/516/oj/eng.

Eurostat. 2021. "Eurostat's Tabular Approach to Exhaustiveness. Guidelines". *5th Meeting of the GNI Expert Group*, 21-22 April 2021. Luxembourg: Eurostat. https://circabc.europa.eu/ui/group/7eb29b7b-33b0-4c9f-851b-e370277bb9e5/library/388028ed-d4dc-4a5f-8909-09128a55ca05/details.

Fawcett, T. 2006. "An introduction to ROC analysis". *Pattern Recognition Letters*, Volume 27, N. 8: 861-874. https://doi.org/10.1016/j.patrec.2005.10.010.

Fernandes, A. 2022. "The non-observed economy in national accounts. "To Be or Not to Be" On The Agenda of National Accountants, Policymakers and Academics". *Ku Leuven Working Paper*. https://hiva.kuleuven.be/en/news/docs/working-paper-the-non-observed-economy-in-the.pdf.

Franz, A. 1983. “Wie groß ist die “schwarze“ Wirtschaft?”. *Mitteilungsblatt der Österreichischen Statistischen Gesellschaft*, Volume 49, N. 1: 1-6.

Gyomai, G., and P. van de Ven. 2014. “The Non-observed Economy in the System of National Accounts”. *OECD Statistics Brief*, N. 18. Paris, France: OECD.

Istituto Nazionale di Statistica - Istat. 2024. *L’economia non osservata nei conti nazionali*. *Anni 2019-2022*. Statistica Report. Roma, Italia: Istat. https://www.istat.it/comunicato-stampa/economia-non-osservata-nei-conti-nazionali-anni-2019-2022/.

Kumar, R., and A. Indrayan. 2011. “Receiver operating characteristic (ROC) curve for medical researchers”. *Indian Pediatrics*, Volume 48: 277-287. https://doi.org/10.1007/s13312-011-0055-4.

Luzi, O., and R. Monducci. 2016. “The new statistical register “Frame SBS”: overview and perspectives”. *Rivista di Statistica Ufficiale/Review of official statistics*, N. 1/2016: 5-14. Roma, Italy: Istat. https://www.istat.it/en/publication/rivista-di-statistica-ufficiale-review-of-official-statistics-12016/.

Majnik, M., and Z. Bosnić. 2013. “ROC analysis of classifiers in machine learning: A survey”. *Intelligent Data Analysis*, Volume 17, N. 3: 531-558.

Organisation for Economic Co-operation and Development - OECD. 2002. *Measuring the Non-Observed Economy. A handbook*. Statistics. Paris, France: OECD Publications Service. https://www.oecd.org/en/publications/measuring-the-non-observed-economy-a-handbook_9789264175358-en.html.

Sallusti, F. 2025. “Measuring Profit Shifting Using “Resident” Information: The PSM-ROC Method”. *IMF Economic Review*, Volume 73, N. 1: 267-315. https://doi.org/10.1057/s41308-024-00238-x.

Schneider, F.G. 2007. “Shadow Economies and Corruption All Over the World: New Estimates for 145 Countries”. *Economics: The Open-Access, Open-Assessment E-Journal*, Volume 1, N. 2007-9: 1-66. https://doi.org/10.5018/economics-ejournal.ja.2007-9.

Schneider, F.G. 2005. “Shadow economies around the world: what do we really know?” *European Journal of Political Economy*, Volume 21, N. 3: 598-642. https://doi.org/10.1016/j.ejpoleco.2004.10.002.

Schneider, F.G., and D.H. Enste. 2000. "Shadow Economies: Size, Causes, and Consequences". *Journal of Economic Literature*, Volume 38, N. 1: 77-114. https://doi.org/10.1257/jel.38.1.77.

Thomas, J. 1990. "Quantifying the Black Economy: 'Measurement without Theory' Yet Again?" *The Economic Journal*, Volume 109, N. 456: F381-F389. https://www.jstor.org/stable/2566011.

United Nations Economic Commission for Europe - UNECE. 2021. *Non-Observed Economy in EECCA and SEE countries. Survey Results*. Geneva, Switzerland: UNECE. https://unece.org/statistics/documents/2021/03/presentations/non-observed-economy-eecca-and-see-countries-survey.

Warnock, D.G., and C.C. Peck. 2010. "A roadmap for biomarker qualification". *Nature Biotechnology*, Volume 28, N. 5: 444-445. https://doi.org/10.1038/nbt0510-444.

Youden, W.J. 1950. "Index for rating diagnostic tests". *Cancer*, Volume 3, N. 1: 32-35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.

A census-based methodology for classifying agricultural holdings: the case of Italy

Roberto Gismondi[1]

Abstract

The seventh general agricultural census, conducted in 2020, showed a sharp decline in the number of small farms compared with 2010. Modernisation of Italian agriculture is an essential objective to ensure the survival of smaller agricultural holdings and the competitiveness of larger farms. Using census microdata, we propose a methodology to estimate the degree of modernisation of Italian agricultural holdings. The methodology is based on five census indicators that refer to specific strategic farm features rather than to structural characteristics of the farm or the farm manager. The number of modernity indicators possessed by each farm forms the basis of the proposed classification criterion. The results show that in 2020, 4.6% of holdings had at least four modernity indicators, while 57.3% had none or only one. Farms with young farm managers are much more modern than the general average.

Keywords: Classification, innovation, market orientation, modernisation, organic farming, standard output.

DOI:https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2024.03.

1 Roberto Gismondi (gismondi@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The author would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

An *agricultural holding*, or *farm*, is a single unit, both technically and economically, operating under a single management and undertaking agricultural economic activities within the economic territory of the European Union, either as its primary or secondary activity. The holding may also provide other supplementary (non-agricultural) products and services.

The national and international debate on the classification criteria for agricultural holdings is extensive and complex and has intensified, above all, due to the progressive transformation of agricultural units from predominantly family-run entities, primarily aimed at self-consumption, to real market enterprises (Rocchi and Stefani 2005).

Although within the EU, statistical institutes have been called upon for years to produce numerous statistical indicators relating to the structure and production of agricultural holdings, their integrated key of reading is not always easy. In other words, the description of the phenomena often prevails over the analysis of the processes of change underway. For example, it is not easy to answer what the level of modernisation of agricultural holdings could be and what differences can be found on a territorial basis.

The literature on the classification criteria for agricultural holdings is vast and heterogeneous.

The simplest classification criteria are essentially based on a single classificatory variable. For example, in Ministry of Agriculture, Food Sovereignty and Forests (MASAF 2009), the main typological classification of agricultural holdings - proposed by the European Union (EU) - is outlined, with adaptations to the Italian context, based on the technical-economic orientation (OTE)[2], which indicates the agronomic specialisation of the farm. Food and Agriculture Organization (FAO 2022) widely uses the concept of family farming, on the basis of which family farms are managed and operated by a family are predominantly reliant on family capital and labour, including both women's and men's. The percentage of the family workforce in the total employed workforce has to be at least 50%. A recent example of such

2 The OTE of an agricultural and livestock farm is determined by the percentage incidence of the standard production (defined in Section 2) of the farm's various production activities (crops and livestock) compared to its total standard production.

classification, based on census results, is provided in National Agriculture Statistical Service (USDA NASS 2021).

Most of the works, known in the literature, which propose multivariate quantitative analyses, are based on essentially descriptive analyses, which stratify agricultural holdings based on surface and/or livestock classes, administrative territorial divisions and other characteristics of the farm or of the farm manager (Lund 2005; López et al. 2008; Glogovetan et al. 2011; Dimitrova 2021).

The proposals suggested by Kostov and Mc Erlean (2006), Etumnu and Gray (2020), Hloušková and Lekešová (2020) are based on approaches oriented towards the use of more sophisticated classification tools, as cluster analysis or mixture of distributions techniques.

In the well-known work by Arzeni and Sotte (2014), which follows the previous work by Sotte (2006), a typological classification criterion is proposed which, based on the results of the sixth agricultural census of 2010, stratifies Italian agricultural holdings based on the joint analysis of different census variables. The indicators examined are: altimetric area, percentage of self-consumed production, days of work, use of passive contracting, age of the farm manager, presence of other gainful activities, share of direct payments from the EU on revenues, educational qualification of the farm manager, presence of quality production, OTE. This is one of the first works that proposes distinct typological and behavioural profiles of farms. It highlighted how the vast majority of Italian agricultural units are not "businesses" in the strict sense but pseudo-family entities of very limited economic size.

According to an approach closer to that proposed in this work, the methodology for classifying agricultural holdings should be more focused on answering a specific question. For instance, Rocchi et al. (2012) propose an empirical analysis aimed at verifying whether Italian farming households are actually poorer than other non-agricultural households. FAO (2019) defines the methodology to be used in order to identify small-scale farmers (see also Section 3.1), with the main goal of calculating the sustainable development goals 2.3.1 and 2.3.2 and of remarking different capabilities of small-scale and not-small-scale farmers in accessing the right quantity of safe food.

In this context, we asked the following question: is the information collected with the seventh general agricultural census (Section 2) able to evaluate the degree of modernisation of Italian agricultural holdings? Being “modern”, or not being so, implies the definition of modernity in the agricultural field, a concept that is obviously very complex and cannot be assessed only on the basis of what the census itself can capture (De Filippis and Henke 2014). The modernisation of the national agricultural system is an unavoidable process, which today can also be sped up on the basis of the National Recovery and Resilience Plan (PNRR), defined in 2021. As regards interventions relating to the world of agriculture, the Ministry of Agriculture, Food Sovereignty and Forests (MASAF)[3] is supporting measures relating to: development of logistics and improvement of energy efficiency for operators in the agri-food chain, innovation and mechanisation in agricultural holdings, improvements in the agro-irrigation system for better management of water resources.

The 2020 census questions were derived from the information needs that emerged at the EU level, connected to multiple aspects of business management that are not always correlated to the concept of modernisation. However, based on the census, it was possible to collect data relating to the five aspects (Section 3), which can determine, even if approximately, the level of modernity of farms. Broadly speaking, being modern means choosing a management model oriented to the principles of sustainable agriculture, integrated with the surrounding territorial and entrepreneurial context, that can guarantee a minimum level of economic well-being for those who manage the farm and offer additional services beyond the basic agricultural production.

The proposed classification methodology aims to address some limitations of pre-existing methodologies (Gismondi et al. 2023), which often overestimate the number of “small” farms. In this framework, the basic idea is to use indicators referable to specific strategic farm choices and features, and not just to the structural characteristics of the farm (production orientation, altitude, location) or of the farm manager (age, gender, educational qualification), which, however, can be used as post-stratification criteria. Even economic results of farms – which are not observed by agriculture censuses – in some way are consequences of modernity features of farms rather than causes.

Based on simple summary indicators (Section 4), Section 5 illustrates

3 https://www.politicheagricole.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/17911.

the main results. Further details are available post-stratifying the analyses according to particular characteristics of the farm or the farm manager (Section 6), or geographical location (Section 7). The comparison with the 2010 census highlights the growth in the degree of modernisation of Italian agricultural holding (Section 8), while in Section 9 a slightly modified methodology is proposed, which uses a greater number of indicators, which however are not available for previous years. Section 10 contains prospective conclusions. In the following, the terms “agricultural holding” and “farm” will be used with the same meaning.

2. The census of agriculture

The seventh general census of agriculture finds its regulatory basis, at the European level, in Regulation (EU) 2018/1091 of the European Parliament and of the Council of 18 July 2018 relating to integrated statistics on agricultural holdings (European Union Commission 2018). Therefore, the 2020 census was mandatory in each European Union country and fully consistent with recommendations from FAO (2017). The census had the main purpose of updating the now-outdated structural data collected with the 2010 census and enriching the available information assets. The most critical feature was the actual state of farm's activity, in a historical context characterised by the progressive concentration of farms and the consequent reduction in the number of very small agricultural units, whose production is often finalised to self-consumption only. The census referred to 1st October 2020 and included questions on the degree of modernisation and integration of farms with respect to the market. Among all we mention: generational turnover (see, for instance, Proctor and Lucchesi 2012), degree of educational training of the farm manager, relevance of non-EU workers, innovation and digitisation introduced on the farm, irrigation tools, the breakdown of farm revenues, and other gainful activities related to agriculture.

Data collection took place between January and July 2021. Approximately 83% of the units in the census list (which included about 1,7 million units) replied to the questionnaire, and approximately 75% declared to be active farms (the remaining 25% includes holdings that have closed down, were temporarily inactive, are off-target or could not be contacted). Istat is now finalising the data dissemination process. Some preliminary analyses related to the evolution process of Italian farms are available in Henke and Sardone (2022).

All data used for elaborations presented in the next Sections are definitive. The main differences with census data already released (Istat 2022a) are due to two reasons: 1) light revisions concerning a few microdata; 2) common lands have been excluded from elaborations, because some relevant census questions – as those concerned with innovation and multifunctionality – could not be addressed to common lands.

Available data include agricultural land and livestock, but do not include revenues. However, based on census data, Istat calculated the standard output for each active farm. The standard output of an agricultural product (crop or livestock), abbreviated as SO, is the average monetary value of the agricultural output at farm-gate price, in euros per hectare or per head of livestock. There is a regional SO coefficient for each product. The sum of all the SO per hectare of crop and per head of livestock in a farm is a measure of its overall economic size, expressed in euros. The standard output can be used to classify agricultural holdings by type of farming and by economic size.

The census counted 1,133,006 agricultural holdings, including common lands. As a matter of fact, the census outlined the sharp decrease in the number of agricultural holdings between 2010 and 2020 (-30.1%). In particular, farms with SO not exceeding 2,000 euros decreased by 52.4%.

3. The five dimensions

The degree of modernisation of agricultural holdings depends on multiple factors. In this context, five factors have been identified, probably not the only ones among all those that can be listed, but all measurable on the basis of the census survey conducted in Italy. The second and fifth factors were not mandatory based on the EU legislation.

Let's note that the indicators selected are focused on particular managerial strategies and do not concern structural features of the farm (as hectares of surface or geographical localisation) or the farm manager (as gender or age). Many works among those mentioned in Section 1 often mix together managerial strategies and structural features, while we prefer to separate them from a conceptual point of view.

3.1 Economic dimension

The economic dimension is a basic indicator for each agricultural holding (Davis et al. 2013). The basic rationale is that each farmer has the right to ensure food security for himself and his household. Even though the census of agriculture did not pick up any economic data (the IFS Regulation does not require any economic information on farms), census data have been used in order to calculate the above-mentioned standard output (SO).

The standard output takes into account land and livestock, but does not consider other sources of income, such as EU subsidies and other gainful activities. The SO (CREA 2023) is a proxy of the true (but unknown) economic revenues of farms.

The economic dimension of farmers is fundamental in the framework of the FAO Sustainable Development Goal 2.3: by 2030, double the agricultural productivity and revenues of small-scale food producers. Even though, according to FAO (2019; 2022), small-scale food producers should be identified according to the combination of the three dimensions given by agricultural land, livestock and net revenues, Gismondi (2023) showed that very similar results could be obtained using the SO in place of the three above-mentioned indicators.

Each modern farm must have a yearly SO larger than a given threshold. Of course, the threshold may be determined in different ways. In this context, we followed the same approach used in the Farm Accountancy Data Network (FADN) context, better known as RICA in Italy. The yearly FADN survey – managed by CREA – does not represent the entire universe of farms active in a given territory, but only those that, due to their economic size, can be considered professional and market-oriented. In Italy, starting from 2014, the minimum threshold for inclusion in the FADN observation field is 8,000 euros in annual standard production value.

So, the first feature taken into account is expressed through the binary variable, equal to 1 (Yes) if the farm has SO≥8,000 euros and equal to 0 (No) otherwise.

3.2 Market orientation

So, the first feature taken into account is expressed through the binary variable, equal to 1 (Yes) if the farm has SO≥8,000 euros and equal to 0 (No) otherwise.

Farmers' market participation is a vital part of agricultural transformation aimed at economic growth and poverty reduction. Market-oriented farming can generate welfare gains through comparative advantages, benefits from economies of scale due to the emergence of large-scale production and technological advances (Zhang et al. 2021).

Market-orientation in agriculture can be defined as the degree to which resources (land, labour and capital) are allocated to the production process and the exchange or sale of agricultural products. Results from several analyses (for instance, Dos Santos and Ahmad 2020) indicate there are economies of scale associated with farms that sell through local and regional markets.

The last agriculture census asked farmers to indicate the share of the final production used for self-consumption. The farmer could select one of these answers:

1. all the value of the final production (100%);
2. over 50% of the final production value;
3. 50% or less of the final production value;
4. no self-consumption (0%).

Of course, the farms that have selected 2 or 3 cannot be attributed with certainty to the market or non-market category. However, type 3 farms can be considered predominantly market-oriented.

Therefore, the second feature taken into account is expressed through the binary variable, equal to 1 (Yes) if the farm answered 3 or 4 and equal to 0 (No) otherwise.

The methodological upgrade of the classification methodology proposed in Section 9 is based on additional information – external to the census – concerning the market orientation of farms.

3.3 Organic farming

Organic farming can be defined as a method of production that places the highest emphasis on environmental protection and, with regard to livestock production, on animal welfare considerations. It avoids or largely reduces the use of synthetic chemical inputs such as fertilisers, pesticides, additives and medicinal products. The production of genetically modified organisms and their use in animal feed are forbidden. It is a part of a sustainable farming system and a viable alternative to the more traditional approaches to agriculture.

A sustainable food system is at the heart of the European Green Deal. The European Commission has set a target of at least 25% of the EU's agricultural land under organic farming and a significant increase in organic aquaculture by 2030.

As a matter of fact, the area used for organic agricultural production in the EU continues to increase. It expanded from 14,7 million hectares in 2020 to 15,9 million ha in 2021, equivalent to 9.9% of the total utilised agricultural area (UAA) in the EU. In Italy (SINAB 2023), in 2022, organic agricultural areas were 2,35 million hectares, the 18.9% of the whole UAA. More details on data sources on organic farming in Italy are available in Gismondi (2020).

Even though organic farming is not the only dimension able to measure the attention to the environment on the part of farmers, this is an important variable measured by the 2020 census. It indicates the propensity of farmers to radically change their production techniques in order to guarantee sustainable agriculture and preserve food safety for consumers.

The third indicator taken into account is expressed through the binary variable, equal to 1 (Yes) if the farm was organic (crops and/or livestock) and equal to 0 (No) otherwise.

3.4 Other gainful activities (OGAs), or differentiation

Gainful activities of the farm comprise all activities other than the basic agriculture production, having an economic impact on the farm. The census questionnaire took into account other gainful activities directly related to the holding, in which either the resources of the holding (area, buildings, machinery, etcetera) or its products are used in the activity. They include: provision of health, social or educational services; tourism (in particular, agro-tourism activities), accommodation and other leisure activities; handicraft; processing of farm products; production of renewable energy; wood processing (e.g. sawing); aquaculture; contractual work using production means of the holding (agricultural or non-agricultural work); forestry.

OGAs constitute a source of additional income compared to basic agricultural production. The diversification of sources of income is important, especially in the presence of economic shocks or other undesired events, such as climate change, natural disasters or wars (Van der Ploeg et al. 2009). Moreover, OGAs respond to new demand needs and allow the characteristics and traditions of the territory to be valorised.

According to the census results, in 2020 65,126 farms had at least one OGA, the 5.7% of the total number of farms. This percentage increased compared with 2010 (4.7%).

In this context, the fourth feature taken into account is expressed through the binary variable, equal to 1 (Yes) if the farm has at least one OGA and equal to 0 (No) otherwise.

The methodological upgrade of the classification methodology proposed in Section 9 is based on a more detailed use of information on OGAs collected by the census.

3.5 Innovation

Innovation in the agricultural and forestry sector is a broad subject, but in general terms, it can be described as a new idea that proves successful in practice. In other words, the introduction of something new (or renewed, a novel change) which turns into an economic, social or environmental benefit for rural practice. Innovation may be technological, non-technological, organisational, or social, and may be based on new or traditional practices. Moreover, the innovations era is often related to agricultural sustainability (Fontana and Fiorillo 2023).

Support for innovation has been a key priority for EU rural development for many years. The trend towards increasing support for innovation was further reinforced within the Common Agriculture Policy (CAP) 2023-2027. Introducing innovation is a cross-cutting objective that must be integrated into all priorities adopted by Member States in their rural development plans.

The latest agricultural census collected two types of innovation-related information. The first one consists of the answers to the question: “In the last three years (2018-2020), has the farm made investments aimed at innovating the technique or production management?”

The second information source derives from AGEA[4] , which collects data on EU subsidies related to innovation supplied to Italian farms. They concern the following rural development measures:

- quality regimes for agricultural and food products;
- investments in tangible assets;
- aid for starting up entrepreneurial activities for non-agricultural activities in rural areas;
- aid for starting up entrepreneurial activities for the development of small agricultural businesses;
- support for investments in the creation and development of non-agricultural activities;

4 Agenzia per le erogazioni in agricoltura. This body manages all payments from the EU to Italian farmers. The Regulation (EU) 2018/1091 asked the national statistical institutes to link the census with the AGEA database at the microdata level.

- agro-climate-environmental payments;
- biological agriculture;
- Natura 2000 payments and payments related to the Water Framework Directive;
- animal welfare.

The fifth feature taken into account is expressed through a binary variable equal to 1 (Yes) if the farm answered “Yes” to the question on innovation and/ or if the farm received at least one of the EU subsidies listed above, and to 0 (No) otherwise.

4. Analysis and synthetic indicators

The basic idea consists of classifying agricultural holdings based on the number of dimensions they possess, from 0 to 5. Obviously, in this way, two farms can have the same “score” even if they possess modernisation features that are partially or totally different from each other.

For each reference domain, n is the number of agricultural holdings, while n_i is the number of agricultural holdings which have i modernisation features – or i “Yes”, for i=0,1,2,3,4,5. For instance, in 2020, n_1 = 340,744 was the number of farms with one “Yes” only (Table 5.2). Therefore:

$$n = \sum_{i=1}^{5} n_i . \tag{4.1}$$

Moreover, we define:

Number of *high level* agricultural holdings = $n_4 + n_5$ (4.2)

Number of *medium level* agricultural holdings = $n_2 + n_3$ (4.3)

Number of *risk level* agricultural holdings = $n_0 + n_1$. (4.4)

Two important and simple relative indicators are given by:

Share of agricultural holdings at risk = $(n_0 + n_1)/n \cdot 100$ (4.5)

Risk ratio = $(n_0 + n_1)/(n_4 + n_5)$. (4.6)

The risk ratio indicates the number of farms at risk for each "high level" farm. Of course, this indicator should be as low as possible, and it cannot be calculated in case $n_4 + n_5 = 0$. Moreover, we can define:

Number of farms with the particular feature $j = m_j$, $j = 1,2,3,4,5$. (4.7)

Among the n_1 agricultural holdings which have one modernisation feature only, m_{j*} is the number of farms which have *exclusively* that particular feature j. Let's note that while label i indicates the number of "Yes", label j indicates the j-*th* feature, where j=1: economic dimension, j=2: market orientation, j=3: organic farming, j=4: other gainful activities, j=5: innovation. For instance, in 2020, m_1=520,425 was the number of farms with SO≥8,000 euros and m_{1*}=66,496 was the number of farms with SO≥8,000 euros and without any other "Yes" (Table 5.3). We have:

$$n_1 = \sum_{j=1}^{5} m_{j*}. \quad (4.8)$$

Therefore, we can also define:

Exclusive effect of the j-*th* feature $= m_{j*} / m_j \cdot 100$. (4.9)

For instance, in 2020, the economic dimension exclusive effect was 12.8% (66,496/520,425·100). The larger the exclusive effect j, the larger is the relative importance of the j-*th* feature for the overall modernisation framework. In case the five exclusive effects are equal, their standard deviation would be zero. On the other hand, large standard deviations will indicate heterogeneous exclusive effects: the various features do not have the same relative importance as regards the overall modernisation picture of the Italian agriculture.

5. The main results

Using the census data referring to the single farms, Table 5.1 was populated, which counts the number of agricultural holdings that had, in 2020, a certain number of characteristics connected to the degree of modernisation, varying between 0 and 5.

There are 32 different combinations of binary variables equal to one if the farm meets that requirement ("Yes") and equal to zero otherwise ("No"). Given that the final score for each farm is obtained by simply adding the number of "Yes", the procedure implicitly attributes the same weight to each of the five characteristics examined.

The significantly more frequent combinations, which overall concern more than seven out of ten farms, are:

- farms without any modern characteristics (n_0=306,675; 27.1% of the total);
- among farms with only one modern characteristic, the 254,610 market-oriented (22.5%);
- among farms with two modern characteristics, the 247,688 market-oriented and with standard production exceeding 8,000 euros (21.9%).

On the other hand, farms with the three characteristics: organic farming, OGAs and innovation, but with standard production not exceeding 8,000 euros and not (yet) market-oriented, are very rare (only 148).

Table 5.1 - Farms with certain modernisation features (32 combinations). Year 2020

SO > 8,000 euro	Market orientation	Organic farming	With OGAs	Innovation	Number of "Yes"	Number of farms	Number of farms (%)
Yes	Yes	Yes	Yes	Yes	5	6,049	0.5
Yes	Yes	Yes	Yes		4	2,150	0.2
Yes	Yes	Yes		Yes	4	23,880	2.1
Yes	Yes		Yes	Yes	4	17,574	1.6
Yes		Yes	Yes	Yes	4	2,360	0.2
	Yes	Yes	Yes	Yes	4	196	0.0
Yes	Yes	Yes			3	16,778	1.5
Yes	Yes		Yes		3	16,495	1.5
Yes	Yes			Yes	3	86,961	7.7
Yes		Yes	Yes		3	907	0.1
Yes		Yes		Yes	3	6,500	0.6
Yes			Yes	Yes	3	4,333	0.4
	Yes	Yes	Yes		3	243	0.0
	Yes	Yes		Yes	3	1,414	0.1
	Yes		Yes	Yes	3	1,432	0.1
		Yes	Yes	Yes	3	148	0.0
Yes	Yes				2	247,688	21.9
Yes		Yes			2	5,863	0.5
Yes			Yes		2	4,882	0.4
Yes				Yes	2	11,509	1.0
	Yes	Yes			2	5,154	0.5
	Yes		Yes		2	4,499	0.4
	Yes			Yes	2	14,059	1.2
		Yes	Yes		2	182	0.0
		Yes		Yes	2	1,074	0.1
			Yes	Yes	2	764	0.1
Yes					1	66,496	5.9
	Yes				1	254,610	22.5
		Yes			1	6,154	0.5
			Yes		1	2,910	0.3
				Yes	1	10,574	0.9
					0	306,675	27.1

Source: Istat, Census of Agriculture

A more concise key of reading is proposed in Table 5.2, which summarises the n_i frequencies defined in Section 4 and proposes a simple reclassification of farms based on the number of modernity characteristics they possess. This is how we pass from high level farms – which have 5 “Yes” (optimal) or 4 “Yes” (almost optimal) – to those medium level – which have 3 “Yes” (medium-high) or 2 “Yes” (medium-low) – and, finally, to the risk level farms

– with 1 “Yes” (almost critical) or 0 “Yes” (critical). In 2020, compared with the 4.6% of high-level farms, 57.3% were found to be at risk level, with the remaining 38.1% medium level.

Table 5.2 - Farms by number of “Yes” (from 5 to 0). Year 2020

Number of "Yes"	Classification	Number of farms	Number of farms (%)
TOTAL	**Whole population**	**1,130,513**	**100.0**
5	Optimal	6,049	0.5
4	Almost optimal	46,160	4.1
3	Medium-high	135,211	12.0
2	Medium-low	295,674	26.2
1	Almost critical	340,744	30.1
0	Critical	306,675	27.1
4+5	High level	52,209	4.6
2+3	Medium level	430,885	38.1
0+1	Risk level	647,419	57.3

Source: Istat, Census of Agriculture

Table 5.3 summarises the m_j frequencies defined in Section 4, which indicate how many farms have or do not have a certain feature, as well as the frequencies m_{j*}, which indicate how many farms present only that specific feature. As was predictable, m_1=520,425 and m_2=699,182 are the predominant characteristics, but while m_{1*}=66,496, m_{2*}=254,610, therefore the market orientation characteristic is much more exclusive than economic size and other characteristics in general, so much so that the exclusive effect reaches the highest level, equal to 36.4%. Organic farming (m_3=79,052) and differentiation (m_4=65,124) appear not yet to be very widespread, while innovation characterises as many as 16.7% of farms (m_5=188,827).

Table 5.3 - Farms with certain modernisation features.Year 2020 (a) (b) (c)

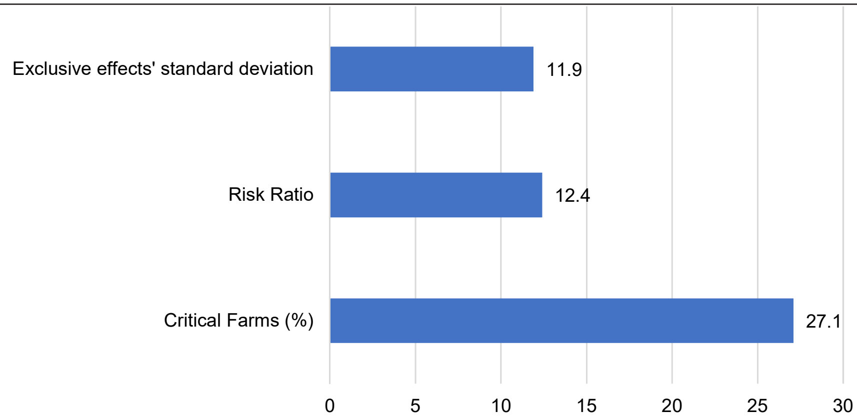
FEATURE	Number of “Yes”	Number of “Yes” (%)	Exclusive “Yes”	Exclusive effect (Ee)
Economic size	520,425	46.0	66,496	12.8
Market orientation	699,182	61.8	254,610	36.4
Organic farming	79,052	7.0	6,154	7.8
Differentiation (OGAs)	65,124	5.8	2,910	4.5
Innovation	188,827	16.7	10,574	5.6

Source: Istat, Census of Agriculture
(a) Number of “Yes”: % ratio between the number of “Yes” and the whole population (1,130,513).
(b) Exclusive “Yes”: number of farms with “Yes” for that particular dimension only.
(c) Exclusive effect: % ratio between exclusive “Yes” and the number of “Yes”.

A very short resume of the main results is provided in Figure 5.1, which shows three main figures:

- the relative share of critical farms (27.1%);
- the risk ratio: for each high-level farm, there are 12.4% farms at risk;
- the standard deviation of exclusive effects: the level 11.9 means that the five modernisation features do not have the same importance, because on average they tend to be exclusive with very different relative frequencies.

Figure 5.1 - Three synthetic indicators. Year 2020 (percentage values)



Source: Istat, Census of Agriculture

The degree of modernity of agricultural holdings is strongly correlated with their dimensional characteristics (Table 5.4). On the one hand, critical farms (with no characteristics of modernity) present on average 1.7 hectares of UAA, 0.1 adult livestock units, 0.17 full-time equivalents, and a relatively low share of non-family workers (8.6%). By contrast, optimal farms have 47.6 hectares, 28 adult livestock units, 3.57 full-time equivalents, and 62.6% of non-family workers.

Table 5.4 - Standard output, utilised agricultural area, adult livestock units and full-time equivalents. Year 2020 (average per farm and percent indicators by number of “Yes” (from 5 to 0))

Number of "Yes"	Standard output (a)	Average per farm			Standard output (% share)	Weight of non-family workers (%)
		Utilised agricultural area (b)	Adult livestock units (c)	Full-time equivalents		
TOTAL	**49,740**	**10.6**	**8.3**	**0.67**	**100.0**	**30.3**
5	233,385	47.6	28.0	3.57	2.5	62.6
4	208,922	37.8	38.5	2.28	17.2	46.0
3	161,168	28.2	31.3	1.68	38.8	35.7
2	66,266	14.0	9.6	0.87	34.8	26.4
1	8,719	4.4	0.9	0.28	5.3	15.1
0	2,673	1.7	0.1	0.17	1.5	8.6

Source: Istat, Census of Agriculture
(a) Euros.
(b) Hectares.
(c) An indicator that summarises in a single number the different varieties of animal species present on the farm.

6. Post-stratification of results

Based on the data collected with the census, it was possible to verify which exogenous factors most influence the presence of farms positioned in the high-level and risk-level clusters. These are post-stratification factors of agricultural holdings belonging to four types: farm characteristics, manager characteristics, kind of production (crops/livestock) and territory (plains/hills/mountain and disadvantaged of non-disadvantaged municipality). Information on disadvantaged municipalities derives from the same data used by Istat (2022b). Their use was made possible through linkage to the census database at the municipal level.

Since the indicator of greatest interest concerns the percentage of farms at risk, in order to evaluate the effect of each post-stratification factor, the factor ratio can be calculated, given by the ratio between the percentage of farms at risk in the particular sub-population identified by post-stratification and the percentage of farms at risk for Italy as a whole (equal to 57.3%). For instance, in the sub-population of 207,085 new farms, the percentage of farms at risk is 50.7%, so the factor ratio is: 50.7/57.3=0.852. In other words, if the farm is new, the probability that it is at risk is on average 14.8% lower than the general average. All results have been summarised in Table 6.1 and in Figure 6.1.

Table 6.1 - Main results using some post-stratification factors. Year 2020

FACTOR	Number of farms	High level	Medium level	Risk level	Risk ratio	Ee Standard deviation (a)
OVERALL	**1,130,513**	**4.6**	**38.1**	**57.3**	**12.4**	**11.9**
Farm characteristics						
New farm	207,085	6.0	43.3	50.7	8.4	9.4
Not a new farm	923,428	4.0	36.5	59.5	14.7	12.5
Management < 3 years	55,124	5.1	47.4	47.6	9.4	9.7
Management ≥ 3 years	1,075,389	4.4	37.3	58.4	13.3	12.0
Family farm	1,114,117	4.4	37.8	57.8	13.1	11.9
Not a family farm	16,396	19.0	62.6	18.4	1.0	6.8
Manager characteristics						
Young (< 40 years)	94,870	12.1	57.7	30.2	2.5	5.4
Not young	1,035,643	3.9	36.3	59.8	15.2	12.5
Male	774,751	5.2	41.2	53.6	10.4	10.6
Female	355,762	3.4	31.4	65.2	19.1	14.8
Basic education	741,704	2.9	35.5	61.5	20.9	13.1
Diploma/degree	388,809	7.8	43.0	49.2	6.3	9.3
Kind of production						
Crops and livestock	186,313	9.7	57.4	32.8	3.4	4.7
Only cultivations	933,405	3.6	34.3	62.0	17.1	13.6
Only livestock	10,795	1.8	30.6	67.6	37.4	6.7
Territory						
Plains	368,957	3.7	43.1	53.2	14.4	12.6
Hills	566,831	4.6	35.3	60.2	13.2	11.9
Mountain	194,725	6.6	36.9	56.6	8.6	10.7
Disadvantaged	260,408	3.1	29.9	67.0	21.3	14.0
Not disadvantaged	870,105	5.1	40.6	54.4	10.7	11.3

Source: Istat, Census of Agriculture
(a) Ee: Exclusive effect.

It is clear that all the factors examined identify behaviours that are different from the general average, confirming how the world of Italian agricultural holdings is extremely heterogeneous internally. However, it is equally clear that some factors discriminate between different behaviours more than others. In particular, the most significant factors are:

- not family farm: the percentage of farms at risk is only 1.0%;
- young managers: the percentage of farms at risk is 2.5%;

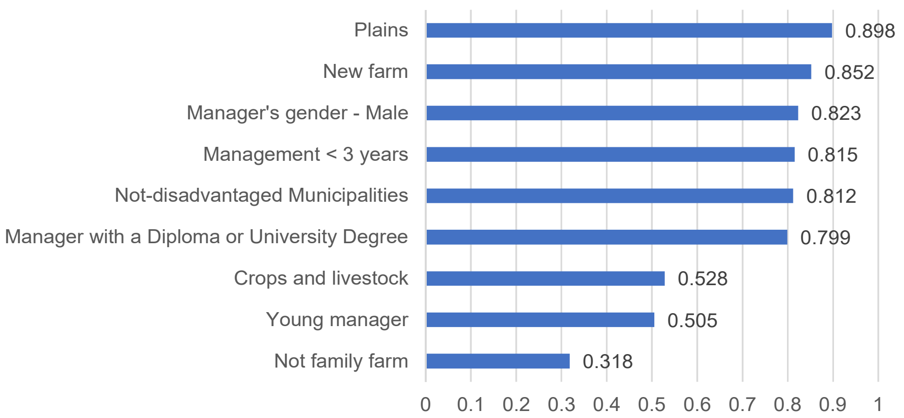
- farms with both crops and livestock: the percentage of farms at risk is 3.4%;
- managers with a diploma or degree: the percentage of farms at risk is 6.3%.

Other factors – although with less intensity than in previous cases – allow us to identify sub-populations of agricultural holdings less at risk than the general average:

- new farms: the percentage of farms at risk is 8.4%;
- farms located in mountain areas: the percentage of farms at risk is 8.6%;
- farms managed by less than 3 years by the farm manager: the percentage of farms at risk is 9.4%.

The lower risk propensity of mountain farms is surprising, at least in part (for those located in flat areas, the share at risk is 14.4%), perhaps due to the fact that the lower accessibility of mountain sites may lead to the need to organise one's own production according to schemes that are basically modern and integrated with the surrounding area. On the other hand, it is undoubtedly comforting to note that the manager's gender does not have much effect on risk propensity, although for female-run holdings this propensity is higher than for male-run ones (19.1% versus 10.4%). Factor ratios in Figure 6.1 summarise what has emerged and are ordered according to an increasing scale, so the first factors are the most discriminating ones.

Figure 6.1 - Factor ratios by post-stratification factor. Year 2020



Source: Istat, Census of Agriculture

7. Geographical details

The geographical breakdown represents one of the most important post-stratification criteria. The persistence of geographical gaps in the degree of evolution of Italian agriculture has already emerged from the first census data released by Istat (2022a). If we focus on the level of modernisation, a somewhat dramatic picture emerges (Table 7.1).

Table 7.1 - Main results by Regions and geographical areas. Year 2020

REGION	Number of farms	High level	Medium level	Risk level	Risk ratio	Ee Standard deviation (a)
Piemonte	50,469	7.2	62.2	30.6	4.3	6.9
Valle d'Aosta/*Vallée d'Aoste*	2,433	6.9	48.1	45.0	6.5	9.9
Lombardia	43,409	7.1	58.2	34.7	4.9	7.0
Liguria	12,416	4.5	41.0	54.5	12.1	6.7
Bolzano/*Bozen*	19,361	20.9	57.2	21.9	1.0	3.7
Trento	13,902	8.6	56.4	35.0	4.1	8.7
Veneto	82,327	4.1	49.9	46.0	11.2	14.7
Friuli-Venezia Giulia	16,217	6.5	45.4	48.1	7.4	14.2
Emilia-Romagna	52,682	10.2	61.1	28.7	2.8	7.2
Toscana	52,3	7.1	36.9	56.1	7.9	8.2
Umbria	27,438	4.9	28.4	66.8	13.7	13.3
Marche	33,669	7.5	33.5	59.0	7.8	14.9
Lazio	62,68	4.6	34.9	60.5	13.0	10.6
Abruzzo	44,293	3.3	31.6	65.1	20.0	11.3
Molise	18,664	1.8	29.6	68.6	38.8	14.3
Campania	79,729	3.1	34.5	62.5	20.2	10.9
Puglia	194,458	2.0	27.3	70.7	35.9	16.1
Basilicata	35,054	2.5	30.8	66.8	26.9	14.1
Calabria	98,278	4.0	21.6	74.4	18.6	14.1
Sicilia	143,72	3.1	36.0	60.9	19.5	14.0
Sardegna	47,014	4.1	51.8	44.1	10.8	6.3
ITALY	**1,130,513**	**4.6**	**38.1**	**57.3**	**12.4**	**11.9**
North-West	108,727	6.8	57.8	35.3	5.5	7.0
North-East	184,489	8.1	54.0	37.9	6.9	10.9
Centre	176,087	5.9	34.2	59.9	10.6	11.1
South	470,476	2.7	28.1	69.2	27.6	14.2
Islands	190,734	3.4	39.9	56.7	17.4	12.1

Source: Istat, Census of Agriculture
(a) Ee: Exclusive effect.

In 2020, in the North-West area, the percentage of farms at risk was only 35.3%, in the South, this share was almost double: 69.2%. The incidence of holdings at risk is much higher than the national average, also in the Centre (59.9%) and in the Islands (56.7%), while in the North-East it is slightly higher than the North-West (37.9%).

Although the North-East area has the highest share of high-level farms, the North-West is the area with the lowest risk ratio (5.5). On the other hand, the risk ratio reaches an extremely high level in the South (27.6), an area in which there is also the lowest relative share of high-level farms (2.7%).

The extreme territorial heterogeneity of the degree of modernisation is further highlighted by regional analyses. In fact, we move from the autonomous province of Bolzano/*Bozen*, with a risk ratio of 1 (only one farm at risk for each high-level farm), to Molise, with a risk ratio of 38.8; the only non-northern regions with a risk ratio lower than the national average are Toscana (7.9), Marche (7.8) and Sardegna (10.8); on the other hand, all the northern regions have a risk ratio lower than the national average.

In the South, modernisation is hindered by the prevalent presence of very small, not very market-oriented agricultural holdings, with few connected and not very innovative activities. In fact (Table 7.2), in the South the percentages of farms, out of the total number of active holdings, with the previous characteristics are respectively equal to 36.2%, 51.1%, 2.4%, and 10.3%, compared to national averages equal to 46.0%, 61.8%, 5.8%, and 16.7%. On the other hand, organic farming is quite widespread in the South, given that the percentage of organic farms is 6.5%, slightly lower than the national average (7.0%) and higher than the share in the North-West (5.2%).

Table 7.2 - Farms with certain modernisation features by geographical areas. Year 2020

		Of which, with certain modernisation features (% on total number of farms)				
AREAS	Number of farms	Economic size	Market orientation	Organic farming	OGAs	Innovation
ITALY	**1,130,513**	**46.0**	**61.8**	**7.0**	**5.8**	**16.7**
North-West	108,727	65.4	78.0	5.2	12.5	26.6
North-East	184,489	61.0	84.8	7.2	10.5	27.6
Centre	176,087	43.7	55.2	9.0	8.7	15.9
South	470,476	36.2	51.1	6.5	2.4	10.3
Islands	190,734	46.9	63.1	7.1	3.0	17.0

Source: Istat, Census of Agriculture

It should also be noted that the average profile of the two largest islands (Sicilia and Sardegna) is quite different from that of the southern regions: although economically small holdings prevail, the agricultural holdings of the Islands are characterised by having market orientation and innovation shares higher than the national average.

The analysis at the municipal scale (Table 7.3) confirms the evidence that emerged at the regional scale. There are 114 municipalities without critical farms (1.4% of municipalities). Overall, 100 municipalities without critical farms are located in the North-West.

In Italy as a whole, 61.0% of the municipalities (4,814 municipalities) have relative shares of risk level farms not exceeding 57.3%, which, as seen in Tables 5.2 and 6.1, is the national average percentage of risk level farms. Consequently, in 3,072 municipalities (39.0%), the relative share of risk level farms exceeds 57.3%. This share reaches 71.5% of the municipalities in the South.

Table 7.3 - Municipalities with certain shares of farms at risk on total farms by geographical areas. Year 2020

AREAS	Number of municipalities	Of which, with a certain share of "Risk level" farms on total farms					
		Percentages on total			Number of critical farms		
		0%	≤57.3%	>57.3%	0%	≤57.3%	>57.3%
ITALY	**7,886**	**1.4**	**61.0**	**39.0**	**114**	**4,814**	**3,072**
North-West	2,982	3.4	79.9	20.1	100	2,382	600
North-East	1,385	0.4	80.8	19.2	6	1,119	266
Centre	971	0.2	39.9	60.1	2	387	584
South	1,781	0.3	28.5	71.5	6	507	1,274
Islands	767	0.0	54.6	45.4	0	419	348

Source: Istat, Census of Agriculture

8. Comparison with 2010

Although agricultural censuses are traditionally aimed at collecting similar and therefore comparable data over time, even if every ten years, each census includes partly or entirely new questions. Therefore, the 2020 census collected data that were not available with the 2010 census. For example, these are the responses provided by agricultural holdings on innovation, which, as described in Section 3.5, represents one of the five dimensions on which the level of modernisation of Italian agricultural holdings in 2020 has been estimated.

Furthermore, the observation fields of the 2020 and 2010 censuses are not exactly the same: while in 2020 fixed minimum size thresholds were used[5], in 2010 these thresholds varied by region. On the other hand, both censuses examined only kitchen gardens managed by farms that also owned other types of crops in addition to kitchen gardens (so both censuses did not include agricultural units that only managed kitchen gardens).

Overall, the 2020 data are substantially comparable with those of 2010. However, given that data on innovation are not available for 2010, in order to be able to measure the changes in the degree of modernisation of Italian farms that occurred over the decade, we opted for the application of a classification methodology slightly different from that described in Sections 3 and 4. In practice, for both 2020 and 2010, the methodology applied differs from that seen so far because the "innovation" dimension has been excluded, starting from the availability of data for each individual farm. The remaining four dimensions (economic size, market orientation, organic farming, and other gainful activities) are measurable for both 2020 and 2010.

The main consequence is that, to allow comparison between 2020 and 2010, the classification of holdings has been as follows:

- high level farms: have 4 "Yes" (optimal) or 3 "Yes" (almost optimal);
- medium level farms: have 2 "Yes";
- risk level farms: have 1 "Yes" (almost critical) or 0 "Yes" (critical).

5 The 2020 census observed agricultural and livestock units with at least: 0.2 hectares of used agricultural surface; or 0.2 hectares planted with vines, greenhouses or mushrooms; or, one adult bovine unit.

Of course, the results for 2020 reported in Table 8.1 and Figure 8.1 are slightly different from those already seen in the previous Sections because they are based on 4 dimensions rather than 5.

With regard to 2020, the exclusion of the innovation dimension led to the growth of the share of both high-level farms and risk-level farms: these shares are equal to 7.7% and 59.7% respectively (Table 8.1), while including innovation in the analysis, the shares were equal to 4.6% and 57.3% (Table 5.2).

These growths derive above all from the fact that in the reduced classification proposed in this Section, the medium level holdings are exclusively those with exactly 2 “Yes”, and which therefore decrease compared to the standard classification analysed in Section 5 (in fact, they go from 38.1% to 32.6%).

However, it is worthwhile to remark that the reduced classification has the sole purpose of allowing comparison with 2010. The most important results tell us that, between 2010 and 2020:

- the percentage of high-level farms grows from 5.1% to 7.7%;
- the percentage of risk-level farms drops from 65.9% to 59.7%;
- in particular, the percentage of critical farms (zero “Yes”) drops from 34.6% to 28.1%;
- the risk ratio (Figure 8.1) drops from 12.8 to 7.8;
- the exclusive effect standard deviation drops from 14.8 to 12.8; so, over the decade, the degree of relative importance of each of the four dimensions examined tends to become more uniform.

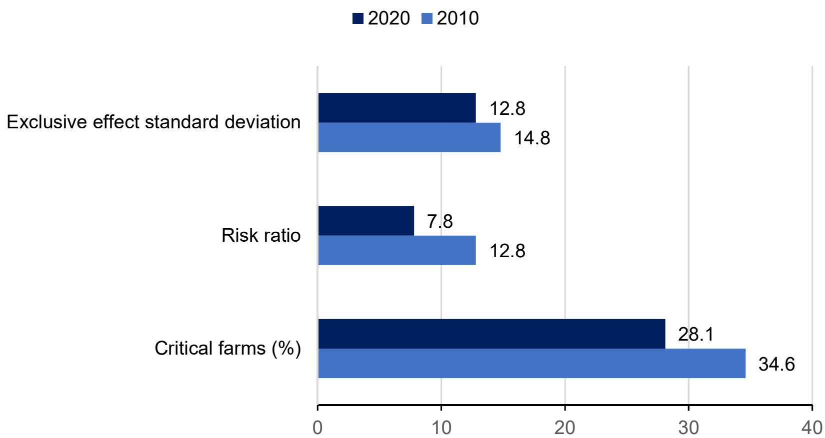
Overall, the results confirm that Italian agriculture is on the path of modernisation, emerging from the essentially rural context that characterised it at least until the 1990s. However, modernisation levels still depend too much on farms’ size and location, according to trends that would be useful to monitor at time intervals closer than a decade.

Table 8.1 - Farms by number of "Yes" (from 4 to 0). Years 2010 and 2020

Number of "Yes"	Classification	Number of farms	Number of farms (%)
Census 2020			
TOTAL	**Whole population**	**1,130,513**	**100.0**
4	Optimal	8,199	0.7
3	Almost optimal	78,433	6.9
2	Medium	369,056	32.6
1	Almost critical	357,576	31.6
0	Critical	317,249	28.1
3+4	High level	86,632	7.7
2	Medium level	369,056	32.6
0+1	Risk level	674,825	59.7
Census 2010			
TOTAL	**Whole population**	**1,620,884**	**100.0**
4	Optimal	6,580	0.4
3	Almost optimal	76,783	4.7
2	Medium	470,119	29.0
1	Almost critical	507,201	31.3
0	Critical	560,201	34.6
3+4	High level	83,363	5.1
2	Medium level	470,119	29.0
0+1	Risk level	1,067,402	65.9

Source: Istat, Census of Agriculture

Figure 8.1 - Three synthetic indicators. Comparison between 2020 and 2010



Source: Istat, Census of Agriculture

9. Methodological upgrade

The methodology proposed in Sections 3 and 4 – the standard methodology – can be further upgraded by introducing changes to the calculation criteria for three of the five indicators of modernity. In fact, again, based on data from the 2020 census, it is possible to evaluate the degree of innovativeness of agricultural holdings also based on specific types of other gainful activities carried out by farms. Furthermore, as seen in Section 3.1, the economic dimension was evaluated using the threshold of 8,000 euros of standard production: this threshold is inevitably subjective – even if it coincides with the threshold that determines the field of observation of the FADN survey – and could be determined according to an alternative logic. Finally, the market orientation was based exclusively on the answers provided by the farms regarding the percentage of production intended for self-consumption, even though it is possible to use a source other than the census to measure this propensity. In order to evaluate how much - and in what direction - the aforementioned methodological changes could modify the results described so far, the entire classification procedure was modified according to the criteria described below.

- Innovation. This feature can be based not only on the question of innovation and the type of subsidies received from the EU, but also on some specific OGAs carried out by the farms. Reference is made to the production of energy from renewable sources (wind, biomass, solar, water, other). Therefore, the innovation dimension can be expressed through the binary variable, equal to 1 (Yes) if the farm answered “Yes” to the question on innovation, and/or if the farm received at least one among the EU subsidies listed above, and/or if the farm produces any kind of renewable energy, and to 0 (No) otherwise. As a consequence, the number of innovative farms should be greather than under the standard methodology.
- Economic dimension. This feature can be expressed as a binary variable equal to 1 (Yes) if the farm has SO≥T euro and 0 (No) otherwise. In the previous Sections, we put T=8,000, while the alternative methodology is founded on the concept of poverty threshold and on the old question about poverty and richness of rural households (Salvioni 2005). This indicator is updated by Istat annually and represents the monetary value, at current prices, of the basket of goods and services considered

essential for each family to avoid serious forms of social exclusion in the reference context (Istat 2023). In this framework, the threshold T depended on the territorial division in which the agricultural holding was located and was based on the standard household composition of 3 adults. On average, the poverty threshold was found to be T=15,990 euros, almost double compared to T=8,000 euros.

- Market orientation. In addition to the census, Istat has another source of information for evaluating the market propensity of Italian agricultural holdings. The ASIA Agriculture Register[6] includes agricultural farms which, based on available administrative sources, are active on the market (Istat 2022c). Therefore, within the alternative methodology, this feature is expressed through the binary variable, equal to 1 if the farm answered 3 or 4 to the census question concerning self-consumption, and/or if the farm is included in the ASIA Agriculture Register, and equal to 0 otherwise.

We note that while the changes introduced regarding innovation and market orientation contribute to increasing the share of farms with modern characteristics, the change relating to the economic dimension works in the opposite direction, as it introduces a higher threshold than that of the standard methodology.

Taking these changes into account, the rest of the methodology remains identical to that described above. The results are summarised in Table 9.1 and in Figure 9.1.

Table 9.1 - Farms by number of “Yes” (from 5 to 0). Year 2020

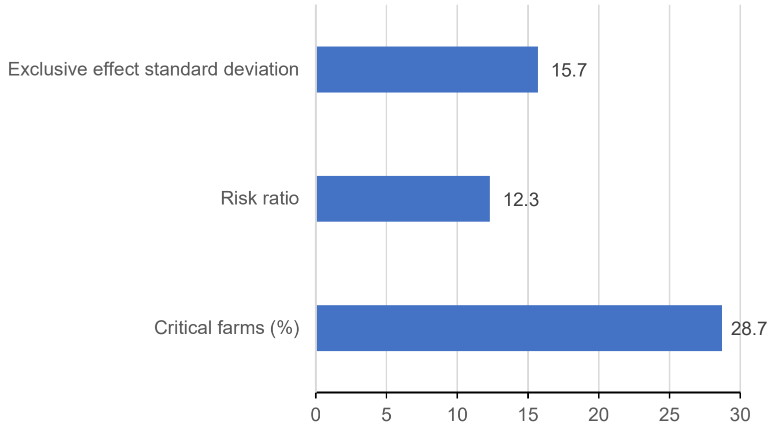
Number of "Yes"	Classification	Number of farms	Number of farms (%)
TOTAL	**Whole population**	**1,130,513**	**100.0**
5	Optimal	11,339	1.0
4	Almost optimal	46,254	4.1
3	Medium-high	124,974	11.1
2	Medium-low	241,612	21.4
1	Almost critical	381,665	33.8
0	Critical	324,669	28.7
4+5	High level	57,593	5.1
2+3	Medium level	366,586	32.4
0+1	Risk level	706,334	62.5

Source: Istat, Census of Agriculture; ASIA Agriculture Register

6 See https://www.istat.it/tavole-di-dati/le-imprese-agricole-in-italia-nel-registro-asia-anno-2020/.

Overall, the alternative methodology leads to results quite similar to those obtained with the standard methodology (Table 5.2). While with the alternative methodology the percentages of high-level, medium-level and risk-level holdings are respectively 5.1%, 32.4%, and 62.5%, these percentages, with the standard procedure, are 4.6%, 38.1%, and 57.3%. While, according to the alternative procedure, the relative weight of farms at risk is higher than in the standard procedure, the risk ratio remains substantially unchanged (12.3% compared with 12.4%).

Figure 9.1 - Three synthetic indicators. Alternative methodology. Year 2020



Source: Istat, Census of Agriculture; ASIA Agriculture Register

In conclusion, it can be stated that the standard methodology defined in Section 3 remains the reference one, even for further applications, and that possible methodological improvements may derive above all from the evaluation of additional dimensions beyond the five examined so far, which cannot be evaluated on the basis of the census data.

10. Conclusions and further steps

The seventh census of agriculture was a huge effort made by all National Statistical Institutes of the European Union. Although most of the data collected through the census were intended to satisfy the requests of the DG-Agri (EU) and Eurostat, the questionnaire used in Italy included additional questions (such as those on self-consumption and innovation).

Using microdata available for each farm, we proposed a methodology to estimate the degree of modernisation of Italian agricultural holdings. The methodology is based on five census indicators concerning specific strategic farm features: economic size, market orientation, recourse to organic farming, other gainful activities beyond the basic agriculture production, and innovation. The number of modernity indicators each farm possesses is the basis for the classification. The main results showed that, in 2020, 4.6% of farms had at least four modernity indicators, while 57.3% had none or only one. Moreover, 1) age of the farm manager, education level and legal form are the most important parameters which determine the degree of modernisation of farms; 2) clear geographical heterogeneities persist in the level of modernisation of holdings, given that all the regions of the South are significantly behind the Centre-North; 3) compared to 2010, the degree of modernisation of agricultural holdings is growing.

One of the limitations of the proposed methodology derives from the fact that each of the five dimensions has the same weight in the synthesis procedure, which allows the classification of agricultural holdings. On the other hand, this aspect is partly only apparent, because based on the analysis of the exclusive effect standard deviations, it follows that market orientation and, to a lesser extent, economic size are much more relevant indicators than the others. That is because often the holdings that present positive performances for these dimensions present positive performances for at least one of the remaining three dimensions.

Furthermore, the use of indicators that are at least in part conceptually interconnected could imply the risk of multicollinearity between the indicators, which, however, is contradicted by the empirical evidence, given that the average linear correlation between the five dimensions was found to be equal to only 0.193.

A further constraint is represented by the impossibility of evaluating other aspects related to the level of modernisation of agricultural holdings, which are not available on the basis of the census. For example, the use of precision agriculture, the level of training of the workforce other than the farm manager, and other environmental protection measures in addition to organic agriculture. For this reason, it is extremely important to increase collaboration among institutions that, in various ways and for different purposes, manage information bases, including administrative ones, relating to agricultural holdings, in order to cross-reference indicators with high informative power at the single-unit level .

An intrinsic limit of the classificatory analysis, as the one proposed, derives, in reality, from the intrinsic structure of Italian agriculture. Some of the dimensions examined, such as organic farming and the presence of other gainful activities, remain infrequent among Italian agricultural holdings, so the relative impact of these indicators on the classification mechanism is limited.

From this perspective, it would be extremely important to replicate the calculations according over time intervals of less than ten years. The convergence process that should reduce the old North-South gap is certainly underway, but it is necessary to monitor its speed and territorial capillarity. We must keep in mind that an agricultural system still so divided into two large blocks, the predominantly modern market holdings and those small and oriented towards self-consumption, is not sustainable anymore (Andreoli and Tellarini 2000; Dos Santos et al. 2020).

No one behind must be the objective to be pursued for Italian agriculture; in the future, it will be extremely important to have data and tools capable of evaluating where we are on the realignment path. From this perspective, the results of the Integrated Farm Statistics 2023 survey, which Istat should disseminate by the end of 2024, will be very useful, because although on a sample and not census basis, they will update the information collected by the 2020 census.

References

Andreoli, M., and V. Tellarini. 2000. “Farm sustainability evaluation: methodology and practice”. *Agriculture, Ecosystems & Environment*, Volume 77, N. 1–2: 43-52. https://doi.org/10.1016/S0167-8809(99)00091-2.

Arzeni, A., e F. Sotte. 2014. “Agricoltura e territorio: dove sono le imprese agricole?”. *QA Rivista dell’Associazione Rossi-Doria*, N. 1/2014: 75-100.

Consiglio per le Ricerche in Economia Agraria - CREA. 2023. *Produzione standard – Standard Output. Definizioni e metodo di calcolo*. Roma, Italia: CREA. https://rica.crea.gov.it/APP/documentazione/?page_id=2153.

Davis, J., P. Caskie, and M. Wallace. 2013. “Promoting structural adjustment in agriculture: The economics of New Entrant Schemes for farmers”. *Food Policy*, N. 40/2013: 90-96.

De Filippis, F., e R. Henke. 2014. “Modernizzazione e multifunzionalità nell’agricoltura del Mezzogiorno”. *QA Rivista dell’Associazione Rossi-Doria*, N. 3/2014: 27-58. https://doi.10.3280/QU2014-003002.

Dimitrova, M. 2021. “Analysis of the typology of agricultural holdings”. *Trakia Journal of Sciences*, N. 19/2021, Suppl. 1: 305-313. http://www.uni-sz.bg/tsj/Volume%2019,%202021,%20Supplement%201,%20Series%20Social%20Sciences/5%20sekcia/formatirani/45.pdf.

Dos Santos, M.J.P.L., and N. Ahmad. 2020. “Sustainability of European agricultural holdings”. *Journal of the Saudi Society of Agricultural Sciences*, N. 19, N. 5: 358-364. https://www.sciencedirect.com/science/article/pii/S1658077X20300242.

Etumnu, C., and A.W. Gray. 2020. “A Clustering Approach to Understanding Farmers’ Success Strategies”. *Journal of Agricultural and Applied Economics*, N. 52/2020: 335-351. http://doi.org/10.1017/aae.2020.4.

European Union Commission. 2018. Regulation (EU) 2018/1091 of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011 (Text with EEA relevance). http://data.europa.eu/eli/reg/2018/1091/oj.

Fontana, E., e V. Fiorillo. 2023. *Agricoltura tra sostenibilità e innovazione*. Milano, Italia: SDA Bocconi - Edagricole - Crédit Agricole.

Food and Agriculture Organization - FAO. 2022. *Family farming. Legal brief. March 2022*. Roma, Italy: FAO. https://www.fao.org/3/cb8227en/cb8227en.pdf.

Food and Agriculture Organization - FAO. 2019. *Methodology for computing and monitoring the Sustainable Development Goal indicators 2.3.1 and 2.3.2*. Working Papers Series, ESS / 18-14. https://www.fao.org/3/ca3043en/CA3043EN.pdf.

Food and Agriculture Organization - FAO. 2017. *World programme for the census of agriculture 2020. Volume 1. Programme, concepts and definitions*. Roma, Italy: FAO. https://www.fao.org/3/i4913e/i4913e.pdf.

Gismondi, R. 2023. "Small-scale farmers revenues: trends in Italy based on different approaches". *ICAS IX: International Conference on Agriculture Statistics*. Washington D.C.: May 17-19.

Gismondi, R., L. De Felici, A. Germani, C. Gnesi, C. Manzi, and P. Nurzia. 2023. "Who are "small" farmers in Italy? Comparing classification criteria using the 2020 agriculture census data". *12th AIEAA Conference - Associazione Italiana di Economia Agraria e Applicata*. Milano: June 22-23.

Gismondi, R. 2020. "L'evoluzione dell'agricoltura biologica in Italia: un'analisi basata sull'integrazione tra fonti". Istat working papers, 2022/4. Roma, Italia: Istat. https://www.istat.it/it/files//2023/01/IWP-4-2022.pdf.

Glogovetan. O.E., B. Samochis, I.G. Dsnils, and C. Crisan. 2011. "Criteria for the classification of agricultural holdings". *Agricultura–Ştiinţă şi practică*, N. 1-2/2011: 65-69.

Henke, R., and R. Sardone. 2022. "The 7th Italian Agricultural Census: new directions and legacies of the past". *Italian Review of Agricultural Economics*, Volume 77, N.3: 67-75. https://doi.org/10.36253/rea-13972.

Hloušková, Z., and M. Lekešová. 2020. "Farm outcomes based on cluster analysis of compound farm evaluation". *Agricultural Economics – Czech*, Volume 66, N. 10: 435-443. https://doi.org/10.17221/273/2020-AGRICECON.

Istituto Nazionale di Statistica - Istat. 2023. *In crescita la povertà assoluta a causa dell'inflazione*. Statistiche report. Roma, Italy: Istat. https://www.istat.it/it/files//2023/10/REPORT-POVERTA-2022.pdf.

Istituto Nazionale di Statistica - Istat. 2022a. *7° Censimento generale dell'agricoltura: primi risultati*. Roma, Italia: Istat. https://www.istat.it/comunicato-stampa/7censimento-generale-dellagricoltura-primi-risultati/.

Istituto Nazionale di Statistica - Istat. 2022b. *La geografia delle aree interne nel 2020: vasti territori tra potenzialità e debolezze*. Statistiche focus, luglio 2022. Roma, Italia: Istat. https://www.istat.it/it/files//2022/07/FOCUS-AREE-INTERNE-2021.pdf.

Istituto Nazionale di Statistica - Istat. 2022c. *Le imprese agricole in Italia nel registro ASIA. Anno 2020*. Nota metodologica. Roma, Italia: Istat. https://www.istat.it/it/files//2022/12/Nota-metodologica-Asia-agricoltura-2020.pdf.

Kostov, P., and S. McErlean. 2006. "Using the mixtures-of-distributions technique for the classification of farms into representative farms". *Agricultural systems*, Volume 88, N. 2-3: 528-537. https://doi.org/10.1016/j.agsy.2005.07.007.

López, C.J. Á., Valiño, J.A.R., and M.M. Pérez. 2008. "Typology, classification and characterization of farms for agricultural production planning". *Spanish Journal of Agricultural Research*, Volume 6, N. 1: 125-136. https://doi.org/10.5424/sjar/2008061-299.

Lund, P.J. 2005. "Aspects of the Definition and Classification of Farms". Conference Paper presented at the *94th EAAE Seminar "From households to firms with independent legal status: the spectrum of institutional units in the development of European agriculture"*. Ashford: 9-10 April. https://doi.org/10.22004/ag.econ.24417.

Ministero delle Politiche Agricole, Alimentari e Forestali - MASAF. 2009. *La tipologia comunitaria di classificazione delle aziende agricole. Regolamento CE n. 1242/2008. Ambito di applicazione, definizioni e principali novità*. Roma, Italia: MASAF. https://www.reterurale.it/flex/cm/pages/ServeAttachment.php/L/IT/D/D.2f9c4f1a8e57a2c302ef/P/BLOB%3AID%3D2447/E/pdf.

United States Department of Agriculture, National Agricultural Statistics Service - USDA NASS. 2021. *Family farms. 2017 Census of agriculture highlights*. Washington, DC, U.S.: USDA NASS. https://www.nass.usda.gov/Publications/Highlights/2021/census-typology.pdf.

Proctor, F., and V. Lucchesi. 2012. *Small-scale farming and youth in an era of rapid rural change*. London, UK, and The Hague, The Netherlands: International Institute for Environment and Development (IIED), and Hivos. https://www.iied.org/sites/default/files/pdfs/migrate/14617IIED.pdf.

Rocchi, B., and G. Stefani. 2005. "The Evolution of Institutional Structure of Italian Agriculture: A Historical Reconstruction through the Debate on Agricultural Holdings Classification". Paper presented at the *94th EAAE Seminar "From households to firms with independent legal status: the spectrum of institutional units in the development of European agriculture"*. Ashford: 9-10 April. https://doi.org/10.22004/ag.econ.24426.

Rocchi, B., Stefani, G., Romano, D., and C. Landi. 2012. "Are Italian farming households actually poorer than other not agricultural households? An empirical analysis". Paper presented at the *1th AIEAA Conference "Towards a Sustainable Bio-economy: Economic Issues and Policy Challenges"*. Trento: 4-5 June. https://doi.org/10.22004/ag.econ.124128.

Salvioni, F. 2005. "Ricchezza e povertà nelle campagne". *Agriregionieuropa*, anno 1, N. 1/2005. https://agriregionieuropa.univpm.it/en/content/article/31/1/ricchezza-e-poverta-nelle-campagne.

Sistema Informativo Nazionale sull'agricoltura Biologica - SINAB. 2023. BIO in cifre 2023. Roma, Italia: SINAB. https://www.sinab.it/sites/default/files/2023-07/BIO%20IN%20CIFRE%202023.pdf.

Sotte, F. 2006. "Imprese e non-imprese nell'agricoltura italiana". Politica Agricola Internazionale - PAGRI, N. 1/2006: 13-30. https://doi.org/10.22004/ag.econ.329196.

Van der Ploeg, J.D., Laurent, C., Blondeau, F., and P. Bonnafous. 2009. "Farm diversity, classification schemes and multifunctionality". *Journal of Environmental Management*, N. 90, Suppl. 2: 124-131. https://doi.org/10.1016/j.jenvman.2008.11.022.

Zhang, J., Mishra, A.K., and S. Hirsch. 2021. "Market-oriented agriculture and farm performance: Evidence from rural China". Food policy, Volume 100: 102023. https://doi.org/10.1016/j.foodpol.2021.102023.

The *Rivista di statistica ufficiale* publishes peer-reviewed articles dealing with cross-cutting topics: the measurement and understanding of social, demographic, economic, territorial and environmental subjects; the development of information systems and indicators for decision support; the methodological, technological and institutional issues related to the production process of statistical information, relevant to achieving official statistics purposes.

The *Rivista di statistica ufficiale* aims at promoting interactions and exchanges between and among researchers, stakeholders, policy-makers and other users who refer to official and public statistics at different levels, to improve data quality and enhance trust.

The *Rivista di statistica ufficiale* was born in 1992 as a series of monographs titled "*Quaderni di Ricerca Istat*". In 1999, the series was entrusted to an external publisher, changed its name to "*Quaderni di Ricerca - Rivista di Statistica Ufficiale*" and started being published on a four-monthly basis. The current name was assumed from the Issue N. 1/2006, when the Italian National Institute of Statistics – Istat returned to being its publisher.

La Rivista di statistica ufficiale pubblica articoli, valutati da esperti, che trattano argomenti trasversali: la misurazione e la comprensione di temi sociali, demografici, economici, territoriali e ambientali; lo sviluppo di sistemi informativi e di indicatori per il supporto alle decisioni; le questioni metodologiche, tecnologiche e istituzionali relative al processo di produzione dell'informazione statistica, rilevanti per raggiungere gli obiettivi della statistica ufficiale.

La Rivista di statistica ufficiale promuove sinergie e scambi tra ricercatori, stakeholder, policy-maker e altri utenti che fanno riferimento alla statistica ufficiale e pubblica a diversi livelli, al fine di migliorare la qualità dei dati e aumentare la fiducia.

La Rivista di statistica ufficiale nasce nel 1992 come serie di monografie dal titolo "Quaderni di Ricerca Istat". Nel 1999 la collana viene affidata a un editore esterno, cambia nome in "Quaderni di Ricerca - Rivista di Statistica Ufficiale" e diventa quadrimestrale. Il nome attuale è stato scelto a partire dal numero 1/2006, quando l'Istituto Nazionale di Statistica - Istat è tornato a esserne l'editore.

e-ISSN 1972-4829
p-ISSN 1828-1982