

# Improving quality in the estimation of “true” economic performance of entrepreneurs by integrating statistical and administrative data: a new method for measuring under-reporting

Rosalba Filippello<sup>1</sup>, Filippo Oropallo<sup>2</sup>, Augusto Puggioni<sup>3</sup>, Paolo Sassaroli<sup>4</sup>

<sup>1</sup>ISTAT, Rome, Italy; [rosalba.filippello@istat.it](mailto:rosalba.filippello@istat.it)

<sup>2</sup>ISTAT, Rome, Italy; [filippo.oropallo@istat.it](mailto:filippo.oropallo@istat.it)

<sup>3</sup>ISTAT, Rome, Italy; [augusto.puggioni@istat.it](mailto:augusto.puggioni@istat.it)

<sup>4</sup>ISTAT, Rome, Italy; [paolo.sassaroli@istat.it](mailto:paolo.sassaroli@istat.it)

## Abstract

The paper describes the new method developed in National Accounts to measure entrepreneurs under-reporting. A description of the new SBS-Frame, an integrated multi-source database, to measure regular economic performance was done. Starting from it, the stratification process and the model based imputation had permitted to estimate non-observed entrepreneurs performance.

**Keywords:** (1-5 words) administrative data, micro-integration, entrepreneur, under-reporting

## 1. Introduction

The development of an integrated database with different administrative sources (Frame of Structural Business Statistics – SBS Frame) to estimate the economic performance of enterprises, has improved the quality of the SBS and National Accounts estimates (NA), eliminating the sampling error and in particular removing bias estimation due to non-response, because concentrated in specific subsets of the universe (see paragraph 2). Thanks to the new sources of economic data, in fact, the measurement of the "regular" component of economic performance of entrepreneurs has been improved. Notwithstanding the enhancing of the regular component, the problem of measuring "true" economic performance including non-observed result due to voluntary underreporting by entrepreneurs still remained (see paragraph 3). The information available at the micro level allowed, at the recent general revision of national accounts, following the introduction of Regulation ESA 2010, in the year 2014, to

define a new method for estimating under-reporting (see paragraphs 4 and 5). In particular, the paper focuses the under-reporting estimation for marginal and micro entrepreneurs.

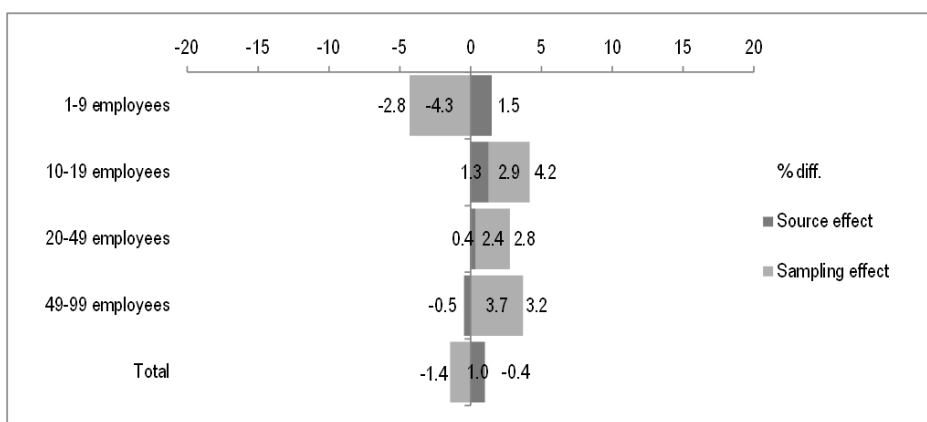
## **2. SBS-Frame**

The SBS Frame is the result of the integration process of different administrative register with Statistical Business Register (SBR) and the Large enterprise census survey (SCI). Since 2011 it replaced the Small-Medium Enterprises (SME) sampling survey for the estimation of the main variables of the economic account and it represents the frame of reference for SBS in Italy. The administrative registers involved in the integration process were the following: a) the financial statements of corporate enterprises; b) Fiscal Authority data: sector study survey and tax return data; c) Social Security data for the labour cost. They were harmonized with the SBS definition and integrated with SBR. It allowed to do a bottom-up estimate of the main SBS variables and had permitted to remove the sampling error that affected final estimate of regular performances of SME. The enhancement in quality was measured in an analysis that was done in the experimental version of the SME Frame (Oropallo, Puggioni et al., 2016).

The quality analysis regarded the following aspects of the integration process: (1) the coverage of the population of reference; (2) the harmonization of variables of the income statement from multiple sources; (3) comparison indicators and distribution of differences; (4) decomposition of differences. The coverage analysis was carried out respecting the importance of different sources. Not all sources have the same informative contents and not all enterprises have the same organization. Financial Statements had a good fit because it covers companies with an ordinary accounting system and this is demonstrated by the comparison indicators with survey data: the distributions of all main variables were similar. They cover about the 16% of the reference population but they represent more than half of its value added. The sole proprietorships and unincorporated enterprises are covered almost all with Sector Studies (about 66%). In this case the harmonization process and the reclassification of the income statement had permitted to estimate the main variables including the value added. Enterprises with a simplified account system, that play a minor role (about 14%), were estimated through Fiscal data. In this case the harmonization process has been more burdensome and the

comparison with survey data was less satisfactory. In general the analysis of the comparisons between administrative data and survey data reveals a good fit for the most important variables and the presence of errors on the side of the survey that affect quality indicators. For the main variables has been confirmed a random distribution of the differences, moreover the analysis of critical domains confirms the absence of systematic errors.

**Fig. 1** - Value added estimate in survey ( $Y_{sme}$ ), from administrative data ( $Y_{admin}$ ) and decomposition of the total difference by size, Year 2010



The decomposition of the differences (Figure 1) showed the prevalence of the sampling effect on the source effect. The first is equal to -1.4 percentage points and the second to +1.0 percentage points, and both contribute to a total difference of -0.4%. This analysis confirm the good fit of administrative data to SBS requirements, with a restrained measurement error as confirmed by the source effect. So the source effect represents a minor part respect to the sampling error that derived only from the SME survey, as a result of the misrepresentation of the final respondents as a consequence of a process of MNAR (Missing Not at Random) in the response process. In fact it has been verified that the estimated performance variable (in this case the Value Added of the enterprise) depend on the response (Oropallo, 2010).

### 3. Entrepreneur “true” economic performance

Non-observed economy (NOE) refers to the set of market economic activities that, for various reasons, escape direct observation and raise problems in the statistical measurement (OECD 2002). The need to take account of such phenomena is particularly relevant in the context of

national accounts, for which the principle of exhaustiveness in the representation of the economic flows is applied. This principle was established in international manuals (SNA and SEC) and is one of the rules supervised by the European Community statistical authorities. The lack of exhaustiveness would make problematic both cross-section both time series international comparisons. According to the SEC2010, introduced in Italy since the reference year 2011, the NOE was calculated as the sum of statistical underground, of economic underground (under-reporting and irregular labour) and of illegal activities that together account in year 2013 for about 206 billion euro, 12.9% of GDP (6.9% is for under-reporting) (ISTAT, 2015).

The estimate of the regular component of the economy was made by the SBS-Frame. Its introduction, greatly decreasing the use of previous sample data, has consequently reduced the incidence of total error in the estimates and the statistical underground.

The shadow economy, which includes the value added hidden by the enterprises due to both the incorrect reporting of accounting data (under-reporting of economic results) and the productive contribution of undeclared work, has been treated by renewing fully existing procedures, exploiting the availability of information guaranteed by the SBS-Frame and a renewed processing of information coming from other surveys. In particular, as described in paragraph 4, a different method to select "anomalous" units (under-reporter businesses), and to assess the extent of this "bias" (the under-reported value added) was associated with each defined subpopulation.

In the Italian productive system the incidence, in terms of number, of the micro enterprises is very high (approximately 80%). For this segment the entrepreneur with both worker and manager role is prevalent. In this context, the economic result had a mixed nature and aims to remunerate the labour input and the business risk. Propensity to under-declaration made through its own deliberate concealment of income and expenses for micro-enterprises is higher. This implies a measurement error of the "true" entrepreneurial economic result, for that a new method has been developed which started with a stratification of the population,

selection of regular vs irregular entrepreneurs and imputation of the under-reporting through statistical and economic behavioral models.

#### **4. Stratification of the reference population**

The reference population for the estimation of under-reported value added is represented by small and medium enterprises (0-99 employees). Large enterprises, from 100 employees up, mostly with foreign subsidiaries, were not considered because they have different tools of tax avoidance in the context of global tax planning. The new estimate of under-reporting procedure involves as a first step a stratification of the population in 5 homogeneous subpopulations according to their structural and economic characteristics (ISTAT, 2014).

*Group 1 - Minimum size units:* it includes very small enterprises or almost informal units, defined as those units in which the entrepreneur's work is completely "replaceable" with that of an employee with the same specialization. They are minimum organization firms, with less than EUR 30,000 turnover, have at most an independent worker, and no employees, and often enjoy tax breaks. The Group 1 has been divided into three further sub-groups.

- *Group 1A* includes: (a) units in marginality economic condition, defined as enterprises with very low final income, less than the fifth percentile of the distribution, and those whose entrepreneurs are less than 30 years old; (b) units with self-employment income and with other income (employed by other enterprises or with presumed retirement income for entrepreneurs above 70 years old).
- *Group 1B* includes: (a) units operating in economic activities that do not require highly specialized skills and training, and do not employ external staff; (b) units whose holders are between 30 and 40 years old, and have low capacity to produce income.
- *Group 1C* includes: units involved in economic activities that require highly specialized skills and training, which use or not use external staff, whose holders are between 40 and 70 years old, with greater ability to produce income.

*Group 2 - Micro units:* it includes enterprises, not belonging to the Group 1, that have a business organization and a reduced production structure, and use labour input different from the entrepreneur's work and that can have a minimum provision of specialized technical assets.

They have less than 10 workers for those operating in manufacturing sectors and less than 6 workers for those operating in the service sectors.

*Group 3 - Organized units:* includes small and medium enterprises that have a more complex organizational and production structure, not belonging to the Group 1 and 2, and that have more than or equal to 10 workers operating in manufacturing sectors and more than or equal to 6 workers operating in service sectors.

*Group 4 - Units belonging to domestic enterprise group:* includes the resident enterprises that are part of groups of companies without foreign relations, with a number of employees below 100.

*Group 5: - Non tractable units:* includes all enterprises hard to control or deal with that, for corporate transformation, for special conditions (i.e. bankruptcy and receivership), for peculiar ownership structure (belonging to groups of companies with foreign relations) and for start-up condition, can be characterized by anomalous situations in the budget declared structure, due to different reasons by deliberate under-reporting; includes also all enterprises with remote possibility of an under-declaration of value added (companies controlled by "government", cooperative societies and private law associations with mutual purposes, enterprises belonging to sectors of economic activities of regulated markets).

**Tab.1** Enterprises, employee and independent workers and value added in SBS-Frame and in subpopulations (years 2011-2013)

Subpopulations	Enterprises						Employee + Independent workers			Value Added		
	2011		2012		2013		2011	2012	2013	2011	2012	2013
	(Num.)	(%)	(Num.)	(%)	(Num.)	(%)	(%)			(%)		
GROUP 1	992,621	22.4	965,437	21.8	938,576	21.5	8.2	8.0	7.9	2.7	2.6	2.5
GROUP 2	2,579,927	58.2	2,615,872	59.1	2,602,553	59.5	42.9	43.8	44.1	34.8	35.6	35.3
GROUP 3	220,008	5.0	219,767	5.0	217,447	5.0	24.7	24.6	24.9	27.9	27.7	28.4
GROUP 4	87,797	2.0	95,427	2.2	98,081	2.2	5.4	5.7	5.9	8.8	9.2	9.4
GROUP 5	552,567	12.5	527,470	11.9	514,793	11.8	18.9	17.9	17.1	25.9	24.9	24.4
<b>Total</b>	<b>4,432,920</b>	<b>100.0</b>	<b>4,423,973</b>	<b>100.0</b>	<b>4,371,450</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

## 5. Selection and estimation of under-reporting

The previous method for the estimation of the under-reporting was based on the comparison between the observed income ( $R_i$ ) and the labour cost paid for the employee ( $Clu_i$ ), under the hypothesis of market perfect competition (Franz, 1985).

The drawback of this approach is that it did not take account of the economic cycle, of market characteristics and of the mixed nature of the remuneration of the entrepreneur job. The necessary condition to the applicability of the method is the presence of independent workers. The new method contemplates the selection of regular entrepreneurs and the estimation of under-reporting taking into account the different nature of the entrepreneur's profit. For each subpopulations specific selection of "regular" and "non-regular" units have been defined.

Selection of group 1, that is minimum size units (marginal or almost-informal entrepreneurs subjected to simplified fiscal and accounting rules), is based on a threshold represented by a shadow income  $\widehat{Clu}_i$ . The shadow income is estimated according to a stratification defined by a statistical analysis using labour cost data from the social security register (INPS-EMENS). The statistical tool used is that of regression trees (CART-Classification And Regression Trees algorithm) that allow to stratify enterprises into homogeneous subsets with respect to the labor cost value (Filippello, Sassaroli, Solari, 2013). The shadow income  $\widehat{Clu}_i$  is calculated, for each identified cluster  $i$  (stratum node level), as the maximum value among the average values for labor cost of blue or white-collar workers, both with full-time contract.

Selection of groups 2, 3 and 4 is carried out through the distribution of the composite indicator of regularity based on the three first latent dimensions (productivity, profitability and cost structure) obtained from a Principal Components Factor Analysis based on a large set of economic indicators. Then the threshold is determined by the ROC (Receiver Operating Characteristic) curve (Carucci, Sallusti, 2014).

After the selection procedures, to measure under-reporting, a profitability explicative model has been developed based on a mark-up assumption (Contini et al., 1992). Under this hypothesis the remuneration of the entrepreneur depends on the variable costs, net of fixed costs which do not vary with the production, and may be written as follow:

$$R_i = \mu'(Int_i + (hd)Clu_i) \quad (1)$$

Where  $R_i$  is the entrepreneur remuneration,  $Int_i$  are the intermediate variable costs divided by independent workers,  $\mu'$  is the mark-up coefficient on variable costs,  $Clu_i$  represents the

remuneration of the employee and the parameter ( $hd$ ) is the variable share of labour cost ( $h$ ) multiplied by the fraction of employees on unpaid workers ( $d$ ). In order to sterilize the relationship from the remuneration of the labour of the entrepreneur, due to the characteristic of the entrepreneur job for this segment of enterprises, the labour cost  $Clu_i$  was subtracted by each side of the equation (1), removing it from the final income, to get a proxy of the entrepreneur profit, which remunerates the risk and the entrepreneur's organizational capacity. The final model specification, where fixed effect have been also introduced, is the following:

$$R_i^* = \beta_1 Int_i + \beta_2 Clu_i + a + \sum_{k,m} d_{k,m} D_{k,m,i} + \varepsilon_i \quad (2)$$

With  $R_i^*$ , the entrepreneur profit, that depends on intermediate variable costs  $Int_i$ , labour costs  $Clu_i$  and fixed effects (constant plus dummies:  $a + \sum_{k,m} d_{k,m} D_{k,m,i}$ ), that allow to restrict the assumption of linearity and permit to control heterogeneity in the entrepreneurs' behaviours. Parameter  $\beta_1$  represents the markup on variable costs ( $\mu'$ ) and  $\beta_2$  is equal to  $(\mu'hd - 1)$ . The model error  $\varepsilon_i$  is supposed to be independent and normally distributed. The method of estimation used a robust regression technique (Yohai, Zamar, 1988) and residual analysis confirmed the assumption of normality of the errors and the absence of heteroscedasticity. The correct estimation of the model parameters gave us the possibility of imputing a true profit  $\hat{R}_i^*$  for entrepreneurs positioned below the regularity threshold. The under-reporting  $UR_i$  is equal to the gap between the estimated income of the entrepreneur (that is, true profit  $\hat{R}_i^*$  plus labour cost  $Clu_i$ ) and the observed (suspected under-reported)  $R_i$  income:  $UR_i = (\hat{R}_i^* + Clu_i) - R_i$ .

The estimation of  $\hat{R}_i^*$  and  $UR_i$  is different for each selected "non-regular" units belonging to the various subsets of the reference population:

Group 1A and Group 5)  $UR_i = 0$  for marginal units in particular or difficult conditions.

Group 1B)  $UR_i = \widehat{Clu}_i - R_i$ , where  $\widehat{Clu}_i$  is the labour shadow income which correspond to the threshold of regularity at a stratum level  $i$  applied for almost informal entrepreneurs (see par. 4) with lower profitability propensity. This approach is similar to the previous (Franz, 1985) with profit  $\hat{R}_i^* = 0$ .



Group 1C)  $UR_i = (\hat{R}_i^* + \widehat{Clu}_i) - R_i$  with  $\hat{R}_i^* = \hat{\beta}_1 Int_i + \hat{\beta}_2 \widehat{Clu}_i + \hat{a} + \sum_{k,m} \hat{d}_{k,m} D_{k,m,i}$  and  $\widehat{Clu}_i$  is the labour shadow income which correspond to the threshold of regularity at a stratum node level  $i$  applied for almost informal entrepreneurs with a higher profitability propensity observed in some specific economic activities (see par. 4). Dummies  $D_{k,m,i}$ , for this subset with aggregate information, are related to  $k=6$  context indicators (Concentration indicator; Capital intensity indicator; Firm demography; Share of irregular jobs; Change in labour input; Geographical area), with  $m=5$  modalities and  $i \cong 13,000$  strata by 2-digit nace and region.

Group 2)  $UR_i = (\hat{R}_i^* + Clu_i) - R_i$  with  $\hat{R}_i^* = \hat{\beta}_1 Int_i + \hat{\beta}_2 Clu_i + \hat{a} + \sum_{k,m} \hat{d}_{k,m} D_{k,m,i}$  and  $Clu_i$  is the average observed compensation for employees. For this set, of about 2,6 million enterprises, parameters were estimated on selected regular units (37% of micro enterprises: about 1 million units) by 2-digit Nace for Centre-North and South and Islands. Dummies  $D_{k,m,i}$ , are 3 size classes, 209 Nace groups at 3-digit level and 21 regions (NUTS-2 level).

As regards groups 3 and 4 the profit model was not applicable to structured enterprises and the measure of the under-reporting is equal to the distance from the regularity threshold in terms of value added (Carucci, Sallusti, 2014).

## 6. Results and conclusions

The distribution for each sub-population of the “true” value added shows that about half (52%) of the total economy revaluation rate regards to the Group of Micro units (group 2) while the 25% is attributable to the Group of Minimum size units (group 1b and 1c), the remaining 23% is due to larger units (groups 3 and 4). The new method has shown significant improvements over the previous one in terms of coverage (in particular, in terms of value added), with much more plausible results from the economic point of view. Ultimately the result of a less biased measure of the "real" value of the economic results of the companies where significant is the contribution of the entrepreneur work has been reached.

## 7. References

- Carucci M., Sallusti F. (2014), Individuazione e correzione della sotto-dichiarazione tramite indicatori economici, internal working paper, ISTAT.
- Contini, B., Revelli, R., Cuneo, S. (1992) “Productivity and imperfect competition” *Journal of Economic Behaviour and Organization* 18(229-248) North Holland.
- Curatolo S., De Giorgi V., Oropallo F., Puggioni A., Siesto G. (2016) “Quality analysis and harmonization issues in the context of the frame of structural business statistics” *Rivista di Statistica Ufficiale* n. 01/2016.
- Filippello R., Sassaroli P., Solari F. (2013), Procedura per il calcolo del “reddito ombra” nel nuovo metodo di rivalutazione del valore aggiunto adottato dall’Istat, internal working paper, ISTAT.
- Franz A. (1985) *Basic Model in Estimates of the hidden economy in Austria on the basis official statistics*, Austrian Central Statistical Office, 1985.
- ISTAT (2014) *Il nuovo approccio alla valutazione della Noe nei Conti Nazionali. Rapporto finale alla Commissione incaricata di ridefinire il sistema di produzione delle stime dell’economia non osservata*. ISTAT-NOE Commission, forthcoming.
- ISTAT (2015) *Non observed economy in national accounts* [www.istat.it/en/archive/175816](http://www.istat.it/en/archive/175816), [www.istat.it/it/archivio/175791](http://www.istat.it/it/archivio/175791).
- OECD (2002) “Measuring the Non-Observed Economy A Handbook”.
- Oropallo F., (2010) “Analisi delle differenze strutturali nella performance economica tra unità rispondenti e unità non rispondenti nella rilevazione dei risultati economici delle piccole e medie imprese (PMI)” *Collana Contributi ISTAT* n. 7/2010.
- Yohai V. J., Zamar R. H. (1988) *High breakdown estimates of regression by means of the minimization of an efficient scale*, *Journal of the American Association*, 83, 406-413.