

# 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

Filippo Oropallo

Augusto Puggioni

Paolo Sassaroli

Istat – Italian National Statistical Institute

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

The new method developed in National Accounts to measure under-reporting is presented

### **What is under-reporting?**

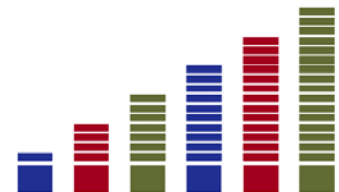
- Deliberate under-reporting of turnover and/or over-reporting of intermediate costs in administrative forms and surveys

### **Why to have a new method?**

- General revision of national accounts, following the introduction of Regulation ESA 2010, in the year 2014
- Availability of SBS-Frame, a census integrated multi-source database

**DATA DRIVEN NEW METHOD, STARTING FROM THE PREVIOUS METHOD, BASED ON SME SAMPLE SURVEY (ABOUT 80,000 UNITS SAMPLED BY 4.5 MILLIONS ENTERPRISES)**

In particular, the presentation focuses **the under-reporting estimation for marginal and micro entrepreneurs**

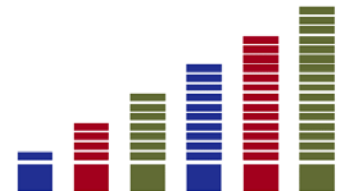




Improving quality in the estimation of “true” economic performance of entrepreneurs by integrating statistical and administrative data: a new method for measuring under-reporting

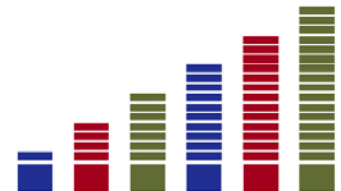
The **SBS Frame** is the result of the integration process of different administrative register (Financial statements, Fiscal data, Social Security data) with Statistical Business Register (SBR) and the Large enterprise census survey (SCI)

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



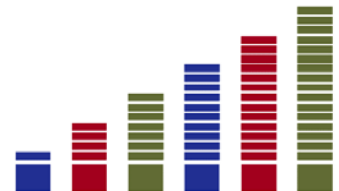
Administrative data were harmonized with the SBS definition and integrated with Statistical Business Register (SBR)

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



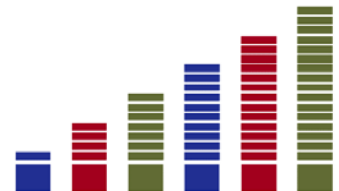
**Non-observed economy (NOE)** refers to the set of market economic activities that, for various reasons, escape direct observation and raise problems in the statistical measurement

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)



The **shadow economy** 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

A new method to select "anomalous" units (under-reporter businesses), and to assess the extent of this "bias" (the under-reported value added) has been defined (larger set of enterprises, economic cycle)



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

## WHY DID WE PUT OUR ATTENTION ON MICRO ENTERPRISES?

Very high incidence (approximately 80%) of *micro enterprises* in the Italian productive system

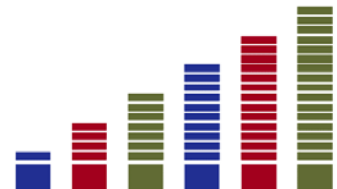
Prevalence for this segment of the **entrepreneur with both worker and manager role**

**Mixed nature of the economic result** that aims to remunerate the labour input and the business risk.

**Higher propensity to under-declaration** made through its own deliberate concealment of income and expenses for micro-enterprises

This implies a **measurement error of the "true" entrepreneurial economic result**, for that a new method has been developed based on three steps:

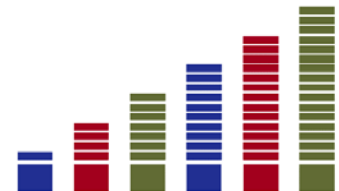
- 1) stratification of the population
- 2) selection of regular vs irregular entrepreneurs
- 3) imputation of the under-reporting through statistical and economic behavioral models



## **Stratification of the reference population (1)**

Small and medium enterprises (0-99 employees)

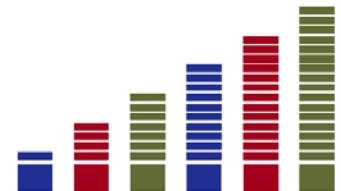
Stratification of the population in 5 homogeneous subpopulations according to their structural and economic characteristics





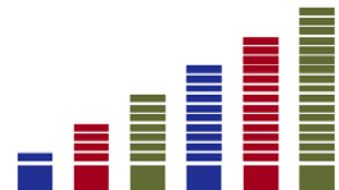
## **Stratification of the reference population (2)**

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



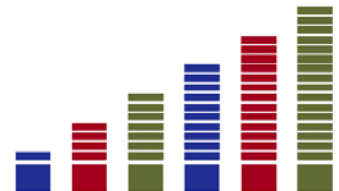
### **Stratification of the reference population (3)**

- Group 1A:
  - (a) units in marginality economic condition
  - (b) units with self-employment income and with other income
- Group 1B:
  - (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: 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



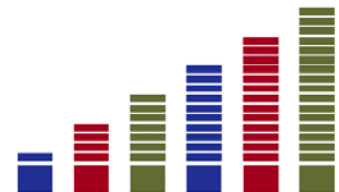
## **Stratification of the reference population (4)**

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.



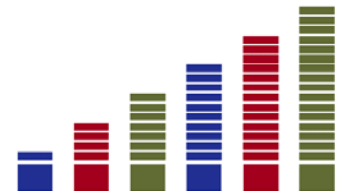
## **Stratification of the reference population (5)**

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.



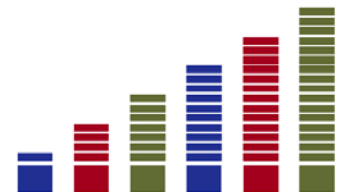
## Stratification of the reference population (6)

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



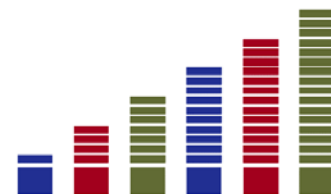
## Stratification of the reference population (7)

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)



## Stratification of the reference population (8)

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>

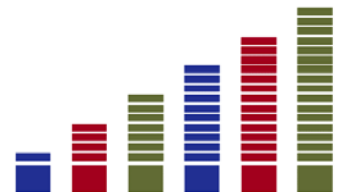


## Selection and estimation of under-reporting (1)

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 (so called *Franz method*)

$Clu_i$  is the “indifference threshold” between self-employment and working as employee

It did not take account of the economic cycle, of market characteristics and of the mixed nature of the remuneration of the entrepreneur job

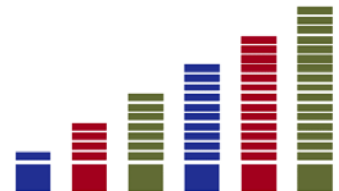




## **Selection and estimation of under-reporting (2)**

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

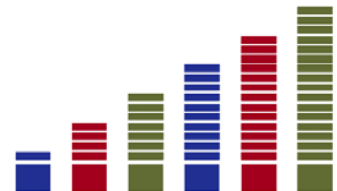


### **Selection and estimation of under-reporting (3)**

Selection of group 1 (minimum size units) is based on a threshold represented by a shadow income  $\widehat{Clu}_i$

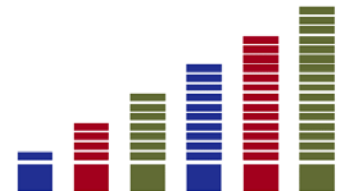
The shadow income is estimated according to a stratification defined by a statistical analysis using labour cost data from the social security register (INPS-EMENS). The statistical tool used is that of regression trees (CART- Classification And Regression Trees algorithm) that allow to stratify enterprises into homogeneous subsets with respect to the labor cost value

The shadow income («opportunity cost»)  $\widehat{Clu}_i$  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 and estimation of under-reporting (4)**

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



### Selection and estimation of under-reporting (5)

After the selection procedures, to measure under-reporting, a profitability explicative model has been developed based on a mark-up assumption. 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:

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

$R_i$  : entrepreneur remuneration

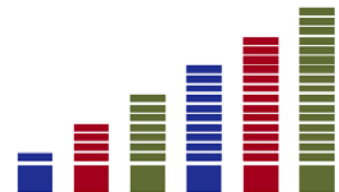
$Int_i$  : intermediate variable costs divided by independent workers

$\mu'$  : mark-up coefficient on variable costs

$Clu_i$  : remuneration of the employee

$hd$  : 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.



## Selection and estimation of under-reporting (5)

The final model specification, where fixed effect have been also introduced, is the following:

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

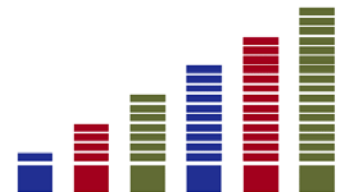
$R_i^*$ : entrepreneur profit

$Int_i$ : intermediate variable costs

$Clu_i$ : labour costs

$a + \sum_{k,m} d_{k,m} D_{k,m,i}$ : fixed effects (it 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

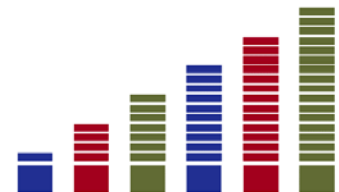


## Selection and estimation of under-reporting (6)

The method of estimation used a robust regression technique 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:  $UR_i = (\hat{R}_i^* + Clu_i) - R_i$



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

### Selection and estimation of under-reporting (7)

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:

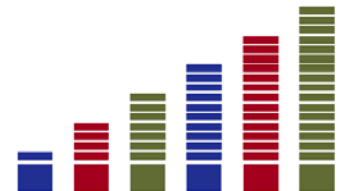
Group 1A and Group 5)  $UR_i = 0$  for marginal units in particular or difficult conditions

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 with lower profitability propensity. This approach is similar to the previous 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

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.

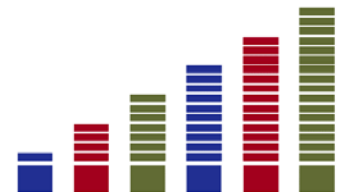


## Selection and estimation of under-reporting (8)

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)

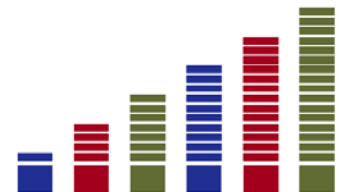
As regards groups 3 and 4 the profit model was not applicable to structured enterprises and the measure of the under-reporting is equal to the distance from the regularity threshold in terms of value added





## Results (1)

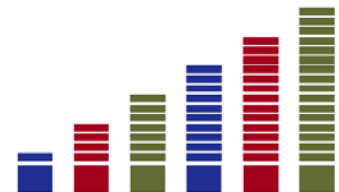
The distribution for each sub-population of the “true” value added shows that about half (52%) of the total economy revaluation rate regards to the Group of Micro units (group 2) while the 25% is attributable to the Group of Minimum size units (group 1b and 1c), the remaining 23% is due to larger units (groups 3 and 4)



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

## Results (2)

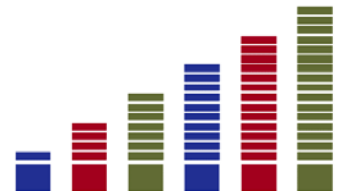
ECONOMIC SECTOR	Value added coverage % (SBS Frame 2011)		Value added revaluation rate%	
	<i>PREVIOUS METHOD</i>	<i>NEW METHOD</i>	<i>PREVIOUS METHOD</i> (SME Survey, Business Accounts)	<i>NEW METHOD</i> (SBS Frame)
MANUFACTURE	54.4%	66.5%	9.1%	8.0%
CONSTRUCTION	71.0%	82.6%	24.5%	17.1%
TRADE, TRANSP., PUBL.SERVICES	70.4%	80.1%	23.3%	26.3%
OTHER SERVICES	63.0%	71.8%	19.8%	24.2%
<b>TOTAL</b>	63.8%	74.1%	18.6%	19.6%



## Conclusions

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





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

# Thank you very much for your attention

Rosalba Filippello ISTAT - [rosalba.filippello@istat.it](mailto:rosalba.filippello@istat.it)

Filippo Oropallo ISTAT - [filippo.oropallo@istat.it](mailto:filippo.oropallo@istat.it)

Augusto Puggioni ISTAT - [augusto.puggioni@istat.it](mailto:augusto.puggioni@istat.it)

Paolo Sassaroli ISTAT - [paolo.sassaroli@istat.it](mailto:paolo.sassaroli@istat.it)

