

# Quality analysis and harmonization issues in the context of “Frame SBS”<sup>1</sup>

Silvana Curatolo, Viviana De Giorgi,  
Filippo Oropallo, Augusto Puggioni, Giampiero Siesto<sup>2</sup>

## Abstract

*The paper describes the results of the quality analysis behind the process of integration and harmonization of administrative data with sample survey of small and medium-sized enterprises (SME). The integration process involves both Structural Business Statistics (SBS) and National Accounts (NA) to reduce the distance from the different estimates. The quality analysis considers a study of the coverage of administrative data, the harmonization of definitions and comparison analysis among administrative and survey data with a subsequent analysis of the distribution of differences in critical domains to verify the absence of systematic errors. The final quality analysis explores separately the impact of the new sources on the SME estimates by decomposing the difference in two parts: (1) the source effect due to the use of administrative data and (2) the sampling effect due to the passage from the sample to census estimations.*

**Keywords:** Micro integration, Administrative sources, Quality indicators

## Sommario

*Il documento descrive i risultati delle analisi di qualità alla base del processo di integrazione e armonizzazione di dati amministrativi con l'indagine sulle piccole e medie imprese (PMI). Il processo di integrazione riguarda sia le statistiche strutturali sulle imprese e sia le stime effettuate nell'ambito dei conti nazionali al fine di ridurre la distanza tra le diverse stime. L'analisi di qualità riguarda lo studio sulla copertura delle fonti amministrative, l'armonizzazione delle definizioni delle variabili e con il confronto tra i dati amministrativi e quelli di indagine con analisi sulle distribuzioni delle differenze in domini critici al fine di verificare l'assenza di errori sistematici. Si mostra, infine, la scomposizione della differenza tra stima campionaria e la nuova stima censuaria in due componenti: (1) “effetto fonte” dovuto dell'utilizzo di dati amministrativi ed (2) “effetto campionario” dovuto al passaggio da stime campionarie a stime censuarie.*

**Parole chiave:** Micro integrazione, dati amministrativi, indicatori di qualità

<sup>1</sup> Although the paper is the result of the work of all the authors, are to be given to: S. Curatolo paragraphs 5.1 and 5.3.1; V. De Giorgi paragraphs 1, 3, 3.1, 3.2, 3.3 and Annex A; F. Oropallo paragraphs 4, 5.2, 6 and 7; A. Puggioni paragraphs 2, 5.3 and Annex B; G. Siesto paragraphs 1, 5.3 and Annex B. The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat.

<sup>2</sup> Istat: [curatolo@istat.it](mailto:curatolo@istat.it), [degorgi@istat.it](mailto:degorgi@istat.it), [oropallo@istat.it](mailto:oropallo@istat.it), [puggioni@istat.it](mailto:puggioni@istat.it), [siesto@istat.it](mailto:siesto@istat.it)

## Introduction

In recent years Istat has intensified the use of administrative sources to support the processes of production of business statistics (Casciano et al. 2012, Luzzi et al. 2013). In particular, the use of financial statement data and economic data from Sector studies and Tax returns, made it possible to estimate the major items of the income statement (turnover, costs, value added, Change in stocks, etc.) for almost all of the statistical units of the Italian Statistical Business Register (Asia).

Thanks to the integration process of all sources, a coverage and a comparison analysis has been done in the context of the sub-working group relating to the quality analysis and preliminary estimation of the frame of SBS (Structural Business Statistics).

Preliminary estimates and comparisons with survey estimates has been done for the main SBS variables and in particular for the variable value added per worker has been evaluated the "sampling effect" (sum of all observations with respect to the weighted sum) respect to the "source effect" (replacing survey data with harmonized administrative data).

### 1. The survey on small and medium enterprises: from sampling design to final estimations

This paragraph describes the production process of the Small and Medium-sized Enterprises survey (SME), by explaining the sample design, the rules for collecting and processing of data, the integration with administrative sources and the estimation procedure used in the SME survey. The paper outlines the domains of estimation for the SBS regulation, the main variables to be transmitted to Eurostat, a short description of SME data production process, the analysis of the available administrative sources having economic information on Italian enterprises and the problems encountered with their use.

The reference population of the SME sample survey is the enterprises with less than 100 persons employed running an industrial, commercial or services activity. This survey, together with the total survey on enterprises with 100 person employed and more (SCI) responds to the requirement of EU Council Regulation on SBS no. 295/2008.

The sampling design is a one stage stratified sampling, with an equal probability of selection for units; the strata are defined by the combination of the modality of the classification of the economic activity Nace Rev.2 (4 digit), size class and administrative region. The sample of the survey referred to the year  $t$  is extracted from Asia register in the year  $t-1$ , in accordance with the domains requested by the SBS regulation. For the reference year 2010, the sample consisted of 100,703 enterprises (about 2.3% of the number of enterprises of Asia).

The data collection of SME survey starts sending a letter to the sampled enterprises (in June of the year  $t+1$ ), which contains the internal code (Asia enterprise code) and the password to access the website of the survey, where they can download the electronic questionnaire and, after compiling it, transmit it via the same website. Since the reference year 2010, the method for collecting questionnaires is totally electronic. The response rate of SME2010 was low, accounting for 38.4%. One of the reasons for this is linked to the complexity of the questionnaire. It consists, in fact, of more than 200 variables (regarding value and cost of production, employment by job category and sex, personnel costs broken

down into different items, external staff with related costs, investments during the year of reference by type of goods, expenditure on environmental protection and other information both quantitative and qualitative) due to the fact it has to be useful not only for complying the SBS regulation but also for national accounts.

The questionnaires, received through the website, are loaded into a RDBMS (Relational Database Management System) and then subjected to an E&I (Editing and Imputation) process that indicates errors and warnings for approximately 23% of the enterprises. Such cases are then solved interactively by data revisers, by comparing survey data with administrative ones or in some cases by contacting the enterprises. The validation process continues until the month of March of the year  $t+2$ .

The phase of integration of non-response begins after the revision, by using the administrative data from financial statements and Sector Studies surveys<sup>3</sup>. The available administrative data represent the macro items of the income statement (according to the scheme of the Fourth EU Directive). The integration of total non-response of SME survey is based on the combined use of administrative data and the donor technique. The phases of the integration process are the following:

1) in order to identify the set of non-response that has to be integrated by administrative data, we compare the validated data file of the SME survey with the files of the theoretical sample and that one used for monitoring arrivals;

2) then, we link the enterprise code, and extract information on the main economic activity of the enterprise and its occupational structure (number of persons employed and employees);

3) afterwards, we use data from financial statements and Sector studies survey to integrate non-response. Obviously, the sources overlap and therefore we need to assign a priority to them, on the basis of information on data quality, consistency between administrative and statistical definitions and the number of available items. The integrated file (no duplication of enterprise codes) will contain all the variables available from the administrative sources together with the structural information from Asia (Nace and employment);

4) total non-response is integrated by identifying for each non-respondent a donor with a similar economic profile (economic activity, size, administrative region). In case we cannot identify the donor enterprise for some strata, the procedure collapse the strata characteristics. The integration process is launched in two steps: first, for the enterprises with employees by selecting donors with personnel costs and, second, for enterprises without employees. Once the group of donors is identified for each stratum, a casual extraction is used to integrate the data of the non-respondent unit with the data of the donor, weighted on the relationship between the persons employed of the non-respondent and the donor persons employed.

The integration process ends with the replacement of the data calculated as described above with those actually declared by the enterprises in the administrative sources for the items of the income statement, while the subheadings are calculated through relations of composition or on the basis of indicators.

At the end of the non-response integration process, position indicators on the whole data

---

<sup>3</sup> See par. 3.2.

set (respondent and integrated enterprises) are calculated in order to identify potential outliers in the different domains, which could have a significant influence on the final estimates. The coverage of the sample through theoretical integration is 80.3%.

On the respondent units the calibration estimates are calculated to obtain final weights that, under certain assumptions, are corrective of residual non-response and error list. They ensure the respect of equality between the totals of the population (number of enterprises and persons employed) and the sample estimates.

The estimator of the total  $Y_{(D)}$  referred to the domain  $D$  is

$$\tilde{Y}_{(D)} = \sum_{k \in s_r} w_k y_k I_k(D)$$

where  $s_r$  is the set of respondent units (respondent and imputed);  $k$  is the unit index,  $w_k$  is the final weight,  $y_k$  is the observed (or imputed) value of the variables of interest;  $I_k(D)$  equals 1 if the unit  $k$  belongs to domain  $D$ , and 0 otherwise.

The final weight  $w_k$  is obtained as a product of three factors:

$$w_k = d_k \gamma_{1,k} \gamma_{2,k}$$

where:

- $d_k$  is the direct weight (the reciprocal of the inclusion probability) ;
- $\gamma_{1,k}$  is the total non-response correcting factor ;
- $\gamma_{2,k}$  is the "post-stratification" factor.

The SBS data transmission to Eurostat (SME sample survey and SCI survey) takes place in June of year  $t+2$  (18 months from the end of the reference year), as showed in Table 1, by defining the statistical confidentiality through the software Tau-Argus.<sup>4</sup>

The SBS Regulation requests the transmission of all the information estimated by the sample survey on the small and medium enterprise and the total survey on the larger enterprises according to the following domains:

**Table 1 - Domains of estimation of the Structural Business Statistics (SBS) n° 295/2008**

1 - Nace Rev.2 at 4 digits without distinction for size classes of persons employed (SME+SCI)						
2 - Nace Rev.2 at 3 digits by for size classes of persons employed:						
	<i>size classes of persons employed</i>					
	0-1	2-9	0-9	10-19	20-49	50-249 250+
<b>Industry and construction –</b> <i>sections: B,C,D,E,F</i>			SME	SME	SME	SME+SCI SCI
<b>Trade and services –</b> <i>sections: G,H,I,J,K(division 66),L, M,N,P, Q,R,S</i> <i>(divisions 95 and 96)</i>	SME	SME		SME	SME	SME+SCI SCI
3 - Nace Rev.2 at 2 digits by administrative regions without distinction for size classes of persons employed (SME+SCI; industry, construction and service)						
4 - Nace Rev.2 at 3 digits by administrative regions without distinction for size classes of persons employed (SME+SCI; trade)						

<sup>4</sup> Software designed to protect statistical tables.

The SBS variables are required for the annexes 1 (services), 2 (industry), 3 (trade), 4 (construction) and 8 (business services that cover enterprises with 20 and more persons employed running in specific economic activities); the most important variables requested are the number of enterprises, turnover, value of production, gross margin on goods for resale, value added at factor cost, gross operating surplus, total purchases of goods and services, purchases of goods for resale, personnel costs, gross salary, hours worked, gross investment in tangible goods, number of persons employed and number of employees.

## **2. Analysis of SBS and National Accounts definitions and other constraints required for the compilation of National Accounts**

National Accounts (NA) are the framework for the quantitative description of the economic and financial situation of an economic system, its components and the relationships between them established in a given period of time. The transactions carried out by economic agents (so called “institutional units”) in relations with other resident in the economic territory units or with non-residents, as well as the changes that these transactions determine on the level and composition of the stock of assets and real or financial liabilities held by them are the matter of the measure.

The elements that characterize the system, in fact, are:

- Systematic and extensive detail in the description of the economy
- Harmonization of methodologies (comparability over time and space)
- Accuracy and precision of concepts, definitions, classifications and accounting rules
- Consistent, reliable and comparable description in terms of quantity of the economies of the EU countries.

According to the criteria of the European Union only a comprehensive measure of GDP makes such aggregate comparable between different Countries and can be used as an indicator for the calculation of contributions from the Member States to the Union, for the control of Maastricht parameters and the allocation of structural funds.

The search for exhaustiveness is guaranteed in the process of construction of the NA by:

- the compliance of definitions and methods with the rules stated by European system of national and regional accounts in the European Union (ESA);
- the measurement of the different components of the non-observed economy<sup>5</sup> and other forms of under-coverage connected to the quality and reliability of information sources.

The gaps between the SBS data and the final NA estimates are traceable and associated to differences in definitions set by the two different regulations, it is the above mentioned concerning SBS and Regulation (EU) No 549/2013 of 21 May 2013 on the European system of national and regional accounts in the European Union (ESA 2010). The two statistical domains serve different, but not conflicting, purposes: in particular ESA is an integrated representation of the economic system, seen as a system of flows between all the different actors, and then a number of strict rules and constraints must be met to assure the consistency.

<sup>5</sup> See for further details “Measuring the Non-Observed Economy - A Handbook” (OECD, 2002).

The main differences between NA, business accounting and SBS are:

1) the definition of production and value added does not totally correspond, since in NA some items need to be reclassified to property income (costs for financial leasing, interest, rent on land, and so on) some to current transfers (insurance premiums and claims);

2) NA estimates do not include capital gains and losses, that have to be eliminated from the values reported in business accounting;

3) a different valuation is required for the economic aggregates: value of production, value added, intermediate costs: producer prices in NA (i.e. net of subsidies on production and products received and including taxes on production and products paid), versus factor cost in SBS (i.e. including subsidies on products and production received net of taxes on products and production paid);

4) NA require that the accounts of enterprises are consistent with those of other enterprises and units in other institutional sectors. So, taxes and subsidies valued on the basis of accounts of enterprises must be consistent with those received by or paid by general government. In practice, this is not observed and a rule is needed in order to achieve consistency. Normally information from general government is more reliable than that from enterprises, and the data from business accounts is adjusted;

5) NA estimates need to be exhaustive, so that a number of specific adjustments are necessary to take into account underreporting of value added, income in kind supplied free of charge to employees, tips;

6) NA estimate domains are more detailed than in SBS: industries in NA are currently defined by Nace 4-digit and size classes (1-5, 6-9, 10-19, 20-99, 100-249, 250-499, 500 and beyond). Therefore the SBS Frame data need to guarantee the joint significance of data for 4-digits Nace Rev.2 domains and for 3-digits Nace Rev.2 and size class; as the data are also used to compile institutional sectors accounts, the legal form of the unit ("corporation", "not corporation") have to be considered, too.

The choice of the methodology for the construction of the SBS Frame has taken into account the needs of both SBS and National Accounts domains, since, when fully operational, the two systems should be integrated: in other words, the flow of data into the new system must provide an intermediate output that constitutes the common input for the compilation of SBS data on the one side, and of NA on the other. It is important to note, though, that to comply with ESA requirements, the valuation and/or classification of some entries of income statement needs to be modified: this generally involves not the "key-variables", but the elementary items that detail the main entries (e.g. the value of insurance premiums paid, rents paid on land, payments for financial leasing, and so on) and that are not always reported in administrative data. As a consequence, all such details have to be estimated in the SBS Frame, by means of a statistical approach grounded on the results of the sample survey.

In the past different approaches were followed. In the NA context a composite estimator was based on SME survey integrated with the population of companies (financial statements). In the SBS context the SME estimates were based on a calibrator estimator with the imputation of missing responses made by integration of financial statements and Sector studies survey (from Italian Revenue Agency) for unincorporated units and sole proprietorships (80% of the Statistical Business Register).

The availability of the new database has totally overhauled the previous approach in the general revision of NA for the transition to ESA 2010<sup>6</sup>: the total amount of the economic aggregates, and value added in particular, as reported in the SBS Frame has been incorporated as such in NA, to represent the observed part of the economic flows. As a consequence, in principle NA and SBS data for SMEs only diverge for some conceptual differences<sup>7</sup>. In general SBS definitions are consistent with business accounting, while national accounting differs on a number of points, because it has a different objective: national accounting aims to describe within a coherent framework all the activities of a country and not just those of an enterprise or group of enterprises. This objective of a consistent picture across all entities in an economy and their relationship with the rest of the world brings constraints which do not affect business accounting.

The revised approach to adjust value added in NA for the under-declaration of income of SMEs because of fraud<sup>8</sup> calls for supplementary information, not needed for SBS (enterprise income as well as other items of the balance sheet), that have to be included in the SBS Frame. The individual information provided by the new data base has also allowed to revise the procedure to adjust value added for the under-declaration of income of SMEs.

On the occasion of the benchmarking revision, the formula to define production and value added from have been modified<sup>9</sup>, for three main reasons:

- a) to achieve a better definition of production and intermediate costs according to the ESA 2010 definitions;
- b) introducing some corrections to the aggregates in accordance with some reservations set by the European Commission on the compilation of Italian Gross national income (GNI);
- c) moving from a definition of production at producer prices to one at basic prices.

A relevant innovation in the calculation of value added is the inclusion in the value of production of a share of revenues not related to the characteristic activity of the enterprise, and registered under the accounting item "Other revenues". Actually, only some of the transactions registered under the item have to be registered as production (for example, insurance claims, capital gains, financial revenues must be excluded), so that a detail of the elementary entries is needed, which is not, however, available nor in administrative nor in the survey data. As a consequence, a statistical procedure have been realized, based on the information available in the Notes to the financial statements of corporations, to elaborate an appropriate structure of the composition of the item "Other revenues" to be used in the SBS Frame to properly sub-divide its value. The development of the SBS Frame represents a substantial innovation in the information system on which NA estimates are grounded and

<sup>6</sup> The transmission to Eurostat of the new series is scheduled by September 2014. The first compilation of the SBS Frame is set for 2011, the year chosen as the benchmarking year for NA. The SBS Frame together with the census-like survey on big enterprises from 100 workers up (SCI) constitute the SBS system data base on business which NA estimates rely on.

<sup>7</sup> In the previous NA methodology per capita values stemming from business surveys are grossed up using estimates of labour input. The new procedure is grounded on an additive approach that distinguishes between observed and non-observed economy. Accounting data as supplied by the SBS, appropriately treated and integrated with other data sources to comply with ESA definitions, provide the total amount the aggregates, limited to the "observed" part of the economy; the results of underground (both concealed to tax authorities and generated by non-registered labour input) and illegal activities are then added.

<sup>8</sup> See "Economia non osservata nei conti nazionali" [www.istat.it/it/archivio/175791](http://www.istat.it/it/archivio/175791).

<sup>9</sup> See "I nuovi conti nazionali in Sec 2010" [www.istat.it/it/archivio/133556](http://www.istat.it/it/archivio/133556).

allows to substantially improves the quality of the accounts.

### 3. Coverage analysis

This part of the paper describes the results of the analyses - reached in the early stages of the working group (January 2013) - on the coverage of the population of the small and medium-sized enterprises sample survey for the reference year 2010<sup>10</sup>, taking into account the administrative data available within Istat. In this context, the coverage is given in terms of number of enterprises, persons employed, employees, and variable VAT-Turnover as they come from Asia in the same reference year. The reference population consists of the enterprises with less than 100 persons employed.

The preliminary analysis has involved the identification of the field of observation in terms of economic activities to be covered (enterprises of industry and non-financial services private sectors) and the study of the available sources of data:

The sources of data identified for the study are the following:

- the Business Register (Asia register);
- the Financial statements of corporate enterprises;
- the Fiscal data (Sector studies and tax return data);
- the Social Security data.

As shown in Table 1, units without coverage represents about 4%. Main issues concerns some domains with a high presence of foreign companies. To guarantee coverage we have adopted two remedies only for *special purpose entities* with more than 3 employees: the first one is to recover the financial statements from other external source; the second one is to reconstruct information by using imputation methods.

**Table 1 - Business Register units by source of administrative data. Year 2010.**

ADMINISTRATIVE SOURCES	Units	%
Financial statements	718,538	16.2
Sector studies	2,930,988	66.0
Tax Return Data	627,671	14,1
No source	166,686	3,8
<b>Total</b>	<b>4,443,883</b>	<b>100.0</b>

#### 3.1 Sources used and domains of coverage analysis

The target population is the same of the reference population used for the extraction of the sample of the survey on SMEs. For the year 2010 the total number of enterprises is 4,443,883 SMEs. For the 2010 SME survey sample, 100,703 unit are extracted: 37,922 are respondents and 42,986 are integrated by administrative sources (income statements and fiscal data).

<sup>10</sup> The year 2010 is the last year available at the date of the analyses.

The administrative data used for these analyses are the financial statements of corporate enterprises, the data of the fiscal survey named Sector studies; the fiscal statements called Modello Unico for corporate enterprises, unincorporated firms and sole proprietorships. In order to complete the analysis of coverage, both administrative and statistical sources (including the SME survey data) have been used. Another fiscal source, the regional tax on productive activities (Irap), has not been used in the early stage.

The coverage analysis is carried out not considering the case of errors or data inconsistency, which means no editing has been done on data. By exploiting the results of Casciano et al. we use a hierarchy of source priority<sup>11</sup>. The analysis is conducted for all domains of reference for SBS, in terms of both the total number of enterprises and the sub-population of firms with employees. The statistical classification of economic activities used is Nace Rev. 2.0, and the domains are the following:

- SBS reference domain:
  - class of economic activity (4-digit Nace Rev.2)
  - group of economic activity (3-digit Nace Rev.2) by size-class of persons employed
  - division of economic activity (2-digit Nace Rev.2) by size-class of persons employed
  - division of economic activity (2-digit Nace Rev.2) by administrative region
- other (non-SBS) domains:
  - division of economic activity (2-digit Nace Rev.2) by legal form
  - division of economic activity (2-digit Nace Rev.2) by class of volume of business<sup>12</sup>

The information related to the number of persons employed, region, VAT-Turnover and legal form are derived from Asia of the reference year, not taking into account the information reported by the SME survey respondent<sup>13</sup>.

### 3.2 Description of the sources

#### *a) The Italian statistical business register (Asia)*

Asia is the statistical business register of Italian enterprises, in accordance with the EEC Regulation no. 2186/93 of 22 July 1993 on "Community coordination of the development of business registers for statistical purposes". Asia represents an official source on the structure of the business population and demography that identifies the enterprises, and their statistical variables. As regards the SME sample survey, Asia has the role of the frame list for extracting the SME samples, as for all Istat business survey. It is also a reference to

<sup>11</sup> 1. financial statements of corporate enterprises, income statement; 2. Sector studies, form F; 3. Sector studies, form G; 4. Modello Unico, form PF-RG; 5. Modello Unico, form PF-RE; 6. Modello Unico, form SP-RG; 7. Modello Unico, form SP-RE; 8. Modello Unico, form SC-RS; 9. Modello Unico, form PF-CM; 10. Modello Unico, form PF-RF; 11. Modello Unico, form SP-RF; 11. Modello Unico, form SC-RF. Priorities are subject to continuous revision and updating, according to different selection criteria. One of them is according to the number of usable variables.

<sup>12</sup> Volume of business is the proxy value for turnover stored in Asia.

<sup>13</sup> The size class is calculated according to the Asia "old" computation of persons employed, in this case preferred to the "new" classification since the present analysis is on the reference year 2010, whose calculation method of the persons employed is the "old" one. We believe, however, that the differences are relatively negligible.

update structural information on enterprises (economic activity, persons employed, employees, etc.) and link through the fiscal code all administrative sources available (Istat, 2011, Nota metodologica). In this context the role of Asia is to extract the structural information and link the administrative data.

*b) The financial statements of corporate enterprises*

They are registered at the Italian Chambers of Commerce. The profit and loss account of the financial statement sets the costs against the revenues for the reporting period, and gives the result of economic operation management.

*c) The Sector studies survey*

This survey is conducted since 1990s by the Ministry of Economy and Finance and represents the tool the Fiscal Authority uses to detect the economic and fiscal parameters for professional activities, self-employed workers and enterprises, in terms of structural as well as economic characteristics, in order to assess their ability to produce income. Sector Studies aim to identify and quantify the voluntary concealment of business performance by comparing the declared economic results with some threshold defined by several direct and indirect indicators. There are some exclusion a non-enforceability principles, the most important one is the threshold for turnover (around 5,2 million euros). Nowadays Sector Studies cover almost all "market" enterprises.

The Sector studies questionnaire is organized into several forms, including those ones relating to economic and financial data used for this study:

- Form F - for enterprises in manufacturing, trade and services sectors;
- Form G - for professional activities.

*d) The Modello Unico*

It is a unified form of tax statement by which the natural and legal persons may submit tax returns, and VAT, Irap and withholding agents tax; the taxpayer obliged either to the compilation of Sector studies survey or its parameters, are also required to fill in the form of Sector studies together with the tax statement.

All income earners in a specific year are obliged to fill in the related form of the Modello Unico, and in general the form of Modello Unico to be filled in is different in accordance with the legal form of the taxpayer. For the purposes of this study, only the forms on the income statement of enterprises and self-employed are usable.

*d.1) The Modello Unico for sole proprietorships (PF)*

Individuals who fill this model are: (i) the natural persons obliged to keep accounting records, (ii) employees, and (iii) generally those who have earned income during the tax period. With regard to income statement and self-employment, the forms used in this analysis are the following:

- form PF-RE - self-employment income from the exercise of arts and professions;
- form PF-RF - income statement of ordinary accounting;
- form PF-RG - income statement of simplified accounting and flat rate schemes;
- form PF-CM - minimum taxpayers.

*d.2) The Modello Unico for unincorporated enterprises (SP)*

Unincorporated firms resident in Italy are required to fill in this form: partnerships; general partnerships and limited partnerships; shipping company; society of fact or irregular (treated as partnerships or partnerships depending on whether or not to run commercial

activity); unincorporated associations among natural persons; marital companies; European economic groups of interest.

With regard to income statement, the related forms used for this analysis are:

- form SP-RE - self-employment income from the exercise of arts and professions;
- form SP-RF - income statement of ordinary accounting;
- form SP-RG - income statement of simplified accounting and flat rate schemes.

#### d.3) *The Modello Unico for corporate enterprises (SC)*

All corporate enterprises resident in Italy are obliged to fill in this form. For this work, we consider:

- form SC-RF – income statement of ordinary accounting;
- form SC-RS - various statements<sup>14</sup>.

#### d.4) *The Irap form*

The Irap form is used to declare the regional tax on productive activities carried out by enterprises. It is filled regardless of the accounting system adopted and is composed of several sub-forms (IQ, IP and IC) in accordance with the different type of the enterprises. This source has been used in the frame process only since the year of reference 2011.

#### e) *The Istat Oros survey*

The Istat short term survey named Oros, which stands for Occupazione (Employment), Retribuzioni (Wages), Oneri Sociali (Other labour cost), aims to produce quarterly information on the evolution (and levels) of gross wage, other labour cost and employment. The Oros survey uses administrative data coming from the Italian Social Security Institute (Inps) for estimating employment, wages and other Personnel costs. The Oros survey is used in this context as a further source to complete the coverage analysis for retrieving information on the cost of labour.

Before moving on coverage and comparisons analyses, it is interesting to examine the Table 2 which contains the information from the SME survey respondents by size classes of persons employed regarding the profit and loss account to which they refer for filling in the questionnaire of the survey<sup>15</sup>.

**Table 2 – Number of respondent enterprises in the SME survey by size classes and profit and loss account used for SME questionnaire. Year 2010**

SIZE CLASSES	Total number of respondents	Chart of accounts	Financial statement (IV directive)	Financial statement (IAS)	Modello Unico	Sector studies	Other	Total
0 - 9	26,348	8,342	9,456	799	8,635	3,415	3,836	34,483
10 - 19	5,636	2,204	3,992	242	698	744	1,145	9,025
20 - 49	3,903	1,562	3,143	175	367	342	974	6,563
50 - 99	1,769	777	1,503	95	162	51	487	3,075
<b>Total</b>	<b>37,656</b>	<b>12,885</b>	<b>18,094</b>	<b>1,311</b>	<b>9,862</b>	<b>4,552</b>	<b>6,442</b>	<b>53,146</b>

Most of enterprises uses financial statement according to the IV Directive or the chart of

<sup>14</sup> At first this source was not available, then the tables in the appendix do not report numbers related to it.

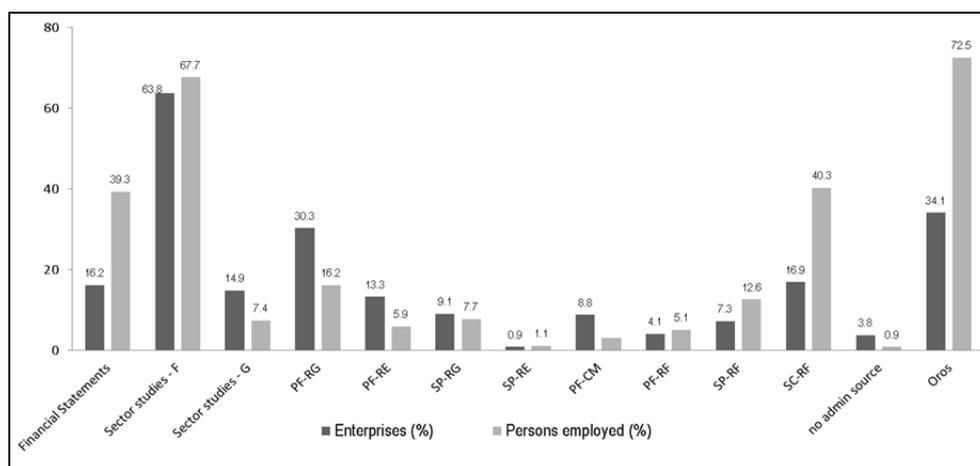
<sup>15</sup> In the questionnaire are allowed multiple responses.

accounts but there are many enterprises that use of Modello Unico (9,862) or Sector Studies (4,552). In the size class 0-9 financial statements in IV Directive, the chart of accounts and fiscal form Modello Unico are used in an almost equivalent manner.

### 3.3 Coverage analysis of the SMEs population

The administrative data of the year 2010 have been preliminarily linked to Asia2010, through the fiscal code, thus identifying the number of enterprises active in the year 2010 for each source. A summary is shown in Figure 1, where if an enterprise is present in more than one source it is counted for each of them.

**Figure 1: Percentages of Asia enterprises and persons employed out of the SMEs reference population by single administrative source. Year 2010.**



As we read in Figure 1, enterprises with financial statements represent 16.2% of the population and 39.3% in terms of employment. These percentages are very different by size classes: they increase according to the size and in the class 50-99 reach percentages of about 87.5% in terms of enterprises and employment.

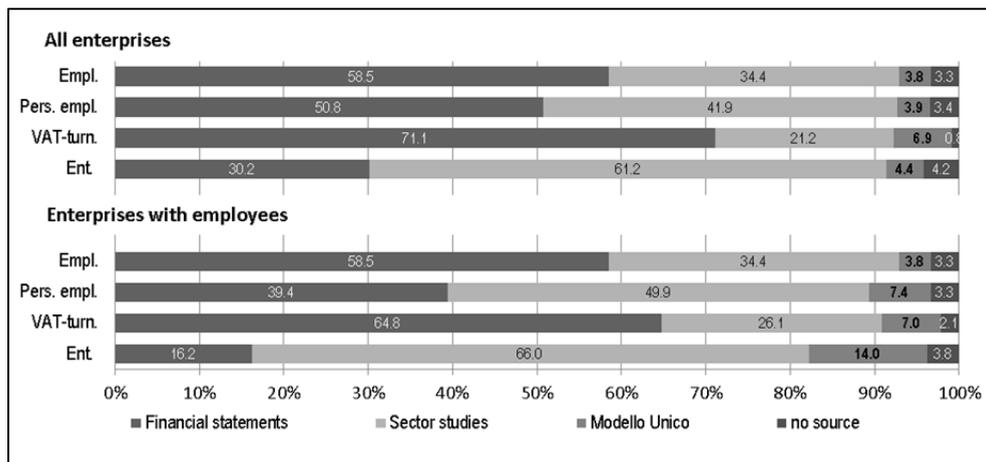
Enterprises that fill in the form F of the Sector studies are more than half and account for more than two-thirds of the employment. The best represented class is that one with 06-09 persons employed, where coverage of enterprises and employment are close to 83%, while the worst represented is the class 50-99 (22-23% of enterprises and employment), mainly due to the turnover threshold.

The form G of the Sector studies represents 14.9% of the enterprises and 7.4% of the employment, percentages in line with the form PF-RE of the Modello Unico.

Very high percentages are covered by the source Oros: 34.1% of the total number of enterprises and 98.8% of those ones with employees<sup>16</sup>.

<sup>16</sup> See Annex A, Table A.1. for more details.

**Figure 2: Percentage of Asia enterprises and persons employed out of the SMEs reference population by overlapped administrative source. Year 2010.**



The sources that covers most of the reference population are the financial statements and the Sector studies. A further significant contribution is given by the form PF-CM of the Modello Unico (8.8% of enterprises) and by the form SC-RF (4.5% of VAT-Turnover)<sup>17</sup>.

The lack of covering is 3.8% in terms of enterprises<sup>18</sup>, 2.1% for VAT-turnover<sup>19</sup> and 3.3% for persons employed. These percentages are more or less the same if calculated for enterprises with employees.

For SBS domains no division, group or class of economic activity by size class and no division by region are uncovered<sup>20</sup>. Poorly covered domains<sup>21</sup> are small or very small ones, and. Also uncovered non-SBS domains are few. In many of these cases data are retrievable from Oros<sup>22</sup>.

<sup>17</sup> See Annex A, Table A.2. for more details.

<sup>18</sup> As mentioned before, the Irap form has been included later in the analysis.

<sup>19</sup> It is worth mentioning that the volume of business, valid for the purposes of VAT, is only a proxy of the turnover, as there are some economic activities exempt from that statement.

<sup>20</sup> Except for a very few cells, in which there is in Asia only one enterprise whose data we cannot retrieve from any source. Similar considerations can be made for enterprises with employees, whose uncovered domains are always represented by 1-2 enterprises in the reference population.

<sup>21</sup> With less than 50% of enterprises of the reference domain.

<sup>22</sup> Uncovered domains involve 131 enterprises, and enterprises in poorly covered domains concern mostly specific legal forms, such as public-economic entities, special entities and public services companies, and fiscal representative entities.

## 4. Harmonization of variables and indicators of comparison

Administrative data has been processed and harmonised in order to replicate the income statement as faithfully as defined in SBS and the SEC regulations. The proposed scheme in Table 3 shows the main variables estimated from administrative sources and then a consistency analysis has been done with the data matched with SMEs survey and the relative theoretical coverage.

**Table 3 – Variables of the profit and loss account re-classified according the value of production**

VARIABLES OF THE PROFITS AND LOSSES ACCOUNT	SMEs Survey	Financial statements	Sector studies	Tax returns data	Coverage
<b>Value of production at costs of factors (PRCF)</b>	<b>PRCF (a)</b>	<b>X</b>	<b>X*</b>	<b>X*</b>	<b>ABC*</b>
+ Turnover	11100	x	x	x	ABC
+ Changes in inventories of finished goods, work in progress and semi-finished	11200	x	x*	x*	ABC*
+ Changes in work in progress	11300	x	x	x	ABC
+ Increase in fixed assets for internal work	11400	x	x		AB
+ Other income (non-financial, non-overtime)	11500	x	x	x	ABC
- Purchases of goods for resale	12103				S
- Change in stockss of goods for resale (opening balance - final amount)	12602				S
<b>Intermediate costs (COSTI_SBS)</b>	<b>COSTI_SBS (b)</b>	<b>X</b>	<b>X*</b>	<b>X*</b>	<b>ABC*</b>
+ Purchases of goods and services	12100+12200	x	x	x	ABC
Purchases - Total	12100	x	x	x	ABC
Services - Total	12200	x	x	x	ABC
+ Charges for the use of third party assets: total	12300	x	x	x	ABC
+ Changes in inventories of raw materials and goods - total	12600	x	x*	x*	ABC*
+ Other operating expenses - total	12900	x	x	x	ABC
- Purchases of goods for resale	12103				S
- Change in stockss of goods for resale (opening balance - final amount)	12602				S
<b>Value Added (VA_SBS= PRCF - COSTI_SBS)</b>	<b>VACF</b>	<b>X</b>	<b>X</b>	<b>X*</b>	<b>ABC</b>
<b>Labour cost (CL)</b>	<b>44000</b>	<b>X</b>	<b>X</b>	<b>X*</b>	<b>ABCD</b>
workers independent	indip				D
workers employees	dip				D
hours worked	orelav				D
<b>Gross Operating Surplus (MOL_SBS=VA_SBS-CL_SBS)</b>	<b>MOL_SBS</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>ABC</b>
Cost of structure (CS_SBS)	12500+12700	X	X	X	ABC
<b>Net Operating Surplus (MON_SBS=MOL_SBS-CS_SBS)</b>	<b>MON_SBS</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>ABC</b>

\* The variable cannot be calculated directly from the administrative data, but can be achieved through a process of estimation.

Sources and coverage:

A → Source: Financial statements (700k/4500k) ~ 16%

B → Source: Sector studies (3000k/4500k) ~ 67%

AB → Source: Financial statements + Sector studies (3400k/4500k) ~ 77%

ABC → Source: Financial statements + Sector studies + Tax return ~ 95%

D → Source: Social Security Data (all enterprises with at least 1 employee ~ 1.300k)

S → Source: Small Medium Enterprises Survey

The calculation of the target variables show, however, problems related to the correct definition of the Value of production, Intermediate costs (which are not subtracted from revenues and cost of goods for resale) and Personnel costs of atypical contractual forms (collaboration etc..) that should be excluded and included in the intermediate costs. The value added is then calculated as the difference between proxies of production value and intermediate costs. The variables turnover, purchases of goods and services, value added, labour cost, gross operating surplus are therefore those directly obtainable from administrative sources.

To validate the information contained in administrative data and to evaluate the consistency of the harmonization process of variable definitions for the main SBS a comparison analysis has been made.

For that purpose, the distributions of the variables in SME survey are compared with the corresponding variable reconstructed from administrative data, calculating the percentage differences compared with the survey estimates.

For each variable, the quality indicators used for comparison are the following:

1. the Kolmogorov-Smirnov index (KS) (similarity of two distributions);
2. the proportion of firms in the range of difference ( $\pm 0\%$ ,  $\pm 2\%$ ,  $\pm 5\%$ );
3. the proportion of value observed in the range of difference ( $\pm 0\%$ ,  $\pm 2\%$ ,  $\pm 5\%$ );
- 4) the average % difference;
- 5) the average value difference;
- 6) the median value difference;
- 7) the interquartile range;
- 8) the coefficient of variation.

## 5. Comparisons between different sources

The aim of this paragraph is to evaluate the consistency of the variables that can be derived from the administrative sources described above with those derived from the SME survey for the reference year 2010<sup>23</sup> through the computation of quality indicators.

### 5.1 Comparison between Financial statements and SME survey

For this analysis the information from 37,920 respondent units to the survey and 742,359 units from Financial statements of corporate enterprises have been linked obtaining 16,602 units useful for the analysis. Therefore, it has been possible to compare the distributions of the variables from SME survey with the corresponding ones reconstructed from the income statement, thus calculating the percentage differences (Table 4).

The comparison between the SME survey data with income statements shows a good fit for many variables. Out of 17 variables comparable, more than half (13) show a distribution more or less similar: this is linked to the characteristics of the survey, which is based on a questionnaire in accordance with the IV accounting directive and to the fact that the editing and imputation phase use the balance sheets to detect and correct the outliers. It should

<sup>23</sup> Only for the Irap form, the reference year is 2011.

however be emphasized that the data of about 20% of the respondent enterprises is subject to a further investigation while 80% of firms are not affected by the process of editing and imputation.

In details, concentrating the analysis on the variables linked to the value of production, the comparison shows a good fit for "Revenue from sales and services", "Change in stocks for work in progress", "Work performed by entity and capitalised"; while the comparisons is definitely not good for the variables "Change in stocks of finished goods" and "Other operating income". Regarding the latter variable, as it can be seen from the table, despite the test is not significant, more than 53% of enterprises have the same value in both sources that represent about 51% of the total in value. However, despite it is a variable that in according to the accounting principles include extraordinary items, such as capital gains, not requested by the SME questionnaire, it is residual and does not affect significantly the estimation of the value added.

**Table 4 - Main indicators of the comparison between the SME survey and income statements variables (16,602 enterprises)**

VARIABLES	KS*	Median diff.	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	0.70	0.0	0.0	0.0	70.5	90.5	93.2	72.3	89.4	91.9
Change in stocks of finished goods (a)	5.75	0.0	0.0	0.0	82.6	84.2	84.5	-	-	-
Change in contract work in progress	0.57	0.0	0.0	0.0	95.2	95.4	95.4	-	-	-
Work performed by entity and capitalised	0.10	0.0	0.0	0.0	98.4	98.6	98.6	78.3	84.6	84.7
Other operating income	7.66	0.0	-0.9	0.0	53.8	58.7	60.4	50.9	61.5	64.2
Purchases of goods	0.95	0.0	0.0	0.0	65,5	79,7	84,6	64,4	84,3	87,9
Purchases of services	0.99	0.0	0.0	0.0	47,3	63,7	70,5	50,4	76,1	82,9
Purchases of goods and services	0.25	0.0	0.0	0.0	47,6	74,8	82,8	51,0	85,6	89,8
Use of third party assets	2.10	0.0	0.0	0.0	74,9	85,0	87,3	73,2	88,3	91,2
Personnel costs	1.67	0.0	0.0	0.0	65,8	82,6	87,3	60,5	82,5	88,6
Wages and salaries	0.29	0.0	-0,1	0.0	46,2	68,0	79,2	34,6	65,3	78,7
Depreciation and amortization	1.49	0.0	0.0	2,2	60,7	71,7	77,8	36,9	57,1	74,2
Change in stocks of raw materials and consumables (b)	3.75	0.0	0.0	0.0	82,1	85,1	85,3	-	-	-
Change in stocks (a-b)	0.66	0.0	0.0	0.0	79,1	83,7	84,1	-	-	-
Other operating expenses	4.34	0.0	0.0	0.0	54,9	61,8	63,6	44,4	51,0	53,2
Value added	1.04	0.0	-1,1	0,3	33,0	58,7	68,0	28,2	63,9	74,4
Gross operating surplus	1.85	0.0	-5,2	0.0	34,2	52,7	58,6	31,5	62,5	69,5

\* Threshold Value 1.6

In terms of costs of production, the only variables that are not significant to test for equal distributions are "Change in stocks of raw materials and consumables" and "Other operating expenses". The KS indicator is slightly negative for the variable "Personnel costs". The variable "Other operating expenses" shows an empirical distribution similar to the corresponding distribution in SME survey. However, even in this case, it could be considered as a residual part of the value added. It represents only 1,7% of the negative components of the value added. Regarding the change in stocks and analysing the single

components there is not a good fit. That depends on errors in filling in the questionnaire for the difficulty of some enterprises to distinguish change in stocks on the side of production from the side of costs. By considering, instead, the variable as a whole "Change in stocks (a-b)", the fit improves so as to ensure that the test is significant.

## 5.2 Comparison between the Sector studies and SME survey

In the Survey we have 37,920 respondent units, of which 14,363 have been matched with Sector studies data. The main results of the comparison analysis on the subset of individuals (sole proprietorships and family businesses, self-employed persons or partnerships) show a positive mean difference of revenues (+0.4%) and also on purchases of goods and services (+0.5%), with a percentage difference for the value added of -1.6%. More detailed analyses at a 3-digit of Nace and company size level of disaggregation do not highlight systematic discrepancies. The highest differences have been reported for the cost items and for units involved in service sectors. It is due to different classifications of costs among different sources, in particular for the item other operating expenses. Other discrepancies have been caused by the mismatch between the economic activity in Asia and in the administrative data (due to the different criteria for identifying the main economic activity). (Table 5).

**Table 5. - Main indicators of the comparison between the SME survey and Sector studies (form F) variables (14,363 enterprises)**

MAIN ECONOMIC VARIABLES	KS*	± 5% (% units)	± 5% (% value)	%diff.	value diff. (000€)	median diff. (000€)	IQR diff. (000€)	CV diff.
Revenues from sales and services	1.0	90.6	94.0	0.4	1.5	0.0	0.0	89.6
Work performed by entity and capitalised	0.4	99.3	47.0	25.5	0.1	0.0	0.0	339.4
Change in contract work in progress	0.5	96.6	522.9	-936.1	0.9	0.0	0.0	88.2
Change in stocks (finished goods, raw materials and goods for resale)	2.1	82.4	31.1	-86.8	-2.5	0.0	0.0	-38.2
Other operating income	9.6	61.5	25.1	-19.5	-1.5	0.0	0.0	-40.1
Purchases of goods (a)	10.1	60.7	76.8	-2.4	-5.1	0.0	2.8	-26.4
Purchases of services (b)	5.7	23.0	26.6	10.8	6.5	0.2	7.8	17.8
Purchases of goods and services (a+b)	1.5	52.4	76.9	0.5	1.4	-0.1	5.3	88.0
Use of third party assets	4.7	80.5	68.1	2.2	0.3	0.0	0.0	48.2
Other operating expenses	15.1	13.1	7.2	-11.1	-1.2	0.0	3.9	-26.8
Personnel costs	4.8	85.1	81.3	1.6	1.0	0.0	0.0	26.5
Depreciation and amortization	3.6	67.8	60.3	-8.3	-1.2	0.0	0.0	-21.9
Value Added	1.1	52.1	48.9	-1.6	-1.9	0.2	5.0	-46.4
Gross operating surplus	1.4	45.6	38.2	-4.7	-2.9	0.1	4.5	-29.7
Net operating surplus	2.1	42.8	36.4	-3.1	-1.5	0.0	5.1	-61.1

\* Threshold Value 1.6

### 5.3 Comparison between "Modello Unico" and SME survey

From the fiscal source minimum taxpayers (Modello Unico PF with CM form), sole proprietorships (Modello Unico PF with RE and RG forms) and partnerships (Modello Unico SP with RE and RG forms) have been considered. Tax return data has been aggregated according to a criterion of proximity to the definitions of the SBS Regulation.

The analysis are conducted at the micro level by comparing distributions and calculating the KS indicator. The indicator measures the similarity of distributions for each couple of variables: the closer it gets to zero the more likely is the hypothesis of similarity. The comparison analysis is carried out without a preliminary editing and imputation process. Some results, then, could be influenced by the presence of outliers.

For the purposes of the analysis, only those enterprises responding to the SME survey (37,656) have been selected, while in the case of the tax data enterprises that fill in the form of the CM minimum taxpayers (393,048 units) or the Modello Unico PF, for the forms RE "income from self-employment" (592,131 units) and RG "determining incomes of enterprise in simplified accounting and flat rate" (1,347,005 units), or the Modello Unico SP (partnerships), for the forms RE (40,229 units) and RG (413,963 units), have been taken into consideration.

Then, it is carried out the linkage by fiscal code present in Asia in order to extract information on economic activity, geographical location and the number of persons employed on which to calculate size classes of persons employed (0-1, 2-3, 4-5, 6-9, 10-19, 20-49 and 50-99). Subsequently, SME survey is linked with the above mentioned sources, obtaining the following results in terms of number of enterprises:

- panel SME - Modello Unico PF (CM): 1,406 enterprises;
- panel SME - Modello Unico PF (RE): 1,835 enterprises;
- panel SME - Modello Unico PF (RG): 6,025 enterprises;
- panel SME - Modello Unico SP (RE): 410 enterprises;
- panel SME - Modello Unico SP (RG): 2,780 enterprises.

The comparison between SME survey and minimum taxpayers data shows a good fit between the "Total of positive components" (CM) and "Revenues from sales and services" (SME) and also the consistency of the variable "Value added" in the two sources is quite satisfactory. For more details, see Annex B, Table B.1.

The comparison between SME survey and tax return data on sole proprietorships (Modello Unico PF), highlights a good fit for "Revenues from sales and services" while for the other items do not seem to be a satisfactory consistency; however, it does not discharge completely on the "Value added". In fact, the value added, according to the KS index, is sufficiently consistent in the two sources, despite the possible presence of costs for project and temporary workers in personnel costs rather than in service costs. That is valid, if considering the tax definitions, but it might not be true that firms matched have actually made use of external staff. For more details, see Annex B, Table B.2 e B.3.

The comparison between SME survey and tax data from Modello Unico SP, RE, and RG highlights the great fit for "Revenues from sales and services" while for the other items, especially for costs, consistency is not always satisfactory, but the "Value added" is sufficiently consistent in the two sources. For more details, see Annex B, Table B.4 e B.5.

The comparison between survey data with Tax return data shows, then, a good level of comparison for the revenue variables and the value added and less for expenses variables, because in the administrative forms a different classification of costs is required. However,

the contribution of such information from Tax return data have to be considered a good proxy of income statements data, that are not available for individuals and partnerships.

In the assessment of administrative sources on the basis of the contribution of information and its consistency with the statistical definitions, a residual role has been assigned to Modello Unico. That finds justification in the fact that tax data contributes to the overall estimate of turnover to about 0.4% (minimum taxpayers CM form) to 0.1% (sole proprietorships RE form) to 0.3% (sole proprietorships RG form) to 0.1% (partnerships RE form) and 0.2% (partnerships RG form). In terms of value added the contributes are 1,2% (minimum taxpayers CM form), 0,3% (sole proprietorships RE form), 0,5% (sole proprietorships RG form), 0,2% (partnerships RE form) and 0,2% (partnerships RG form).

### 5.3.1 Comparison between Irap form and SME survey

The source Irap has been introduced in the process of frame from the year of reference 2011, so in this case the comparisons between the main variables available from the different sources have been based on the data of the reference year 2011.

**Table 6 - Main indicators of the comparison between Irap and SME survey variables (14,461 enterprises)**

VARIABLES	KS*	Median diff.	% units			% value		
			±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	1.64	0.0	65.2	88.4	91.6	73.0	91.6	93.4
Change in stocks of finished goods (a)	40.48	0.0	81.0	82.6	82.7	-	-	-
Changes in contract work in progress	28.39	0.0	94.5	94.6	94.6	-	-	-
Work performed by entity and capitalised	24.66	0.0	98.2	98.3	98.4	76.6	78.1	78.3
Other operating income	19.49	0.0	56.2	62.9	64.3	57.7	72.6	75.0
Purchases of goods	10.13	0.0	61.8	78.6	83.4	67.6	89.5	91.7
Purchases of services	0.63	0.0	38.8	58.4	65.7	44.1	69.6	77.7
Purchases of goods and services	0.43	0.0	38.0	68.6	77.6	51.3	86.3	91.3
Use of third party assets	14.95	0.0	70.3	81.6	84.1	70.8	86.9	89.6
Depreciation and amortization	4.27	0.0	76.6	89.4	91.0	77.6	91.6	93.7
Change stock of raw materials and consumables (b)	33.00	0.0	78.8	81.7	81.9	-	-	-
Other operating expenses	4.88	0.0	45.3	53.9	56.2	43.7	52.2	54.3
Change in stocks (a-b)	28.49	0.0	78.2	82.9	83.3	-	-	-
Value added	0.57	0.0	29.4	56.4	65.9	31.1	63.0	73.2

\* Threshold Value 1.6

The tax is calculated as the difference between the amount of revenues and the amount of the costs. Some costs are not deductible and among these, personnel costs, costs for other workers assimilated to employees (project workers), taxes on the buildings are the most significant ones. The value added calculated from this source should be higher than that one from the statistical and the other administrative sources (SME, income statements, Sector studies and so on). In particular, the Irap source should estimate a lower cost for services (due to the presence of external workers in the other sources) and lower other operating

costs (due to the presence of taxes on buildings in the other sources).

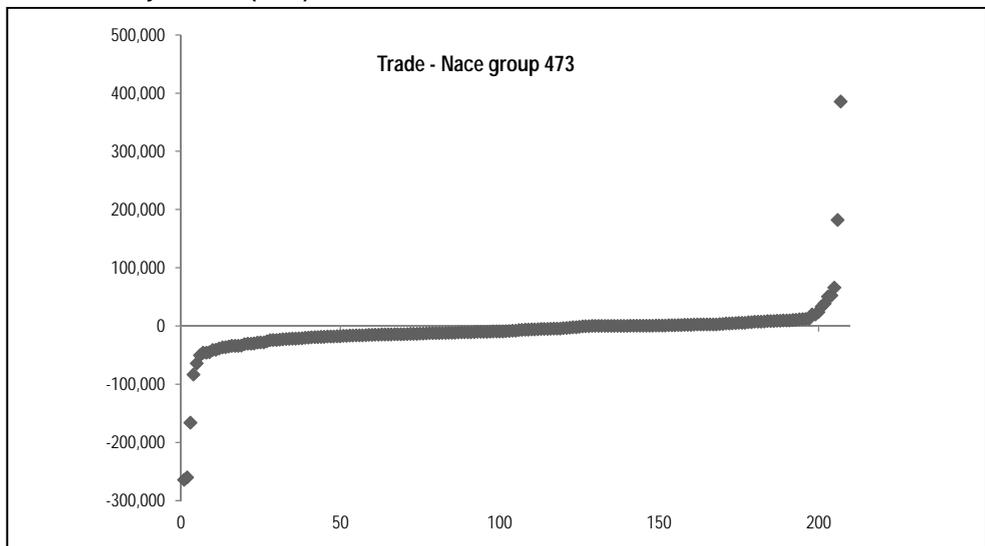
The analysis regards the comparison between the Irap data for the corporate companies (form IC section I) and SME survey data. The comparison involves 14,461 units in both sources and shows a good fit for most variables, except for those ones that include non-deductible costs for Irap such as services costs and other operating costs (Table 6). Several enterprises show the same value in both the sources, although the results improve significantly considering the ranges of variation  $\pm 2\%$ . With regard to the value added in the comparison is better in the range  $\pm 5\%$  that includes 65.9% of enterprises, representing 73.2% of the value.

## 6. Analysis of critical domains

When domains showed relevant and significant differences a deeper analysis has been made. This analysis has been carried out at a group level of Nace by size class and confirms the absence of any systematic discrepancies. The highest differences have been found for the cost variables and relate to the different classification of costs between the different sources examined, in particular for the item other operating costs. Other discrepancies are related to the mismatch between the main economic activity from Asia and the economic activity from the Sector studies source (due to the different criteria for identifying the main economic activity).

As regards the problem of the domains not covered, that is, with a high presence of foreign enterprises, the solution is to exclude from Frame (and Asia) the legal form "1900-special purpose entities without a stable plant" (with less than 3 employees), but only for those with more than 3 employees will need to recover the financial statement data via other sources: Telemaco, Irap Fiscal data.

**Figure 3 – Absolute differences of the value added in Sector studies and survey – Nace group 473 year 2010 (euro)**



The most critical economic activity of the Trade sector (Figure 3) do not show a clear sign of systematic differences. In all cases, the gains or losses depend on a few influential observations that mark in a positive or negative difference in the final evaluation. The test KS for comparing the distributions do not report a significant difference between the distributions of the Value added observed in the two sources. Moreover, the share of the observed value for the companies belonging to the range difference of  $\pm 10\%$  is always greater than 50%.

There is a different classification of revenues for some companies, particularly those that make buying and selling of real estate, so the differences with SMEs relate to the different classification of revenues from sales in other revenues. For firms in the Nace sector 681 (Buying and selling real estate) there are 4 outliers that affect the negative difference. For these companies, revenues should be adjusted for the purchase of buildings to be sold and change in inventories. This information should be retrieved from the survey (the same way as is done for trade firms for goods to resale) or specific information about sales by kind of activity by using additional fiscal data (i.e. form D of the Sector studies).

The transport sector (Nace 494) has a prevalent tail of negative differences (Sector studies < SMEs) and the difference in value added is equal to -8.1% compared to the SMEs (the KS indicators is not significant). There seems to be a problem of non-deductibility of costs from the side of the Sector studies. The cause of the discrepancy concerns the indication of higher operating expenses in the Sector studies (+ 41%), while for the other revenues (-3.5%) prevails SMEs.

The service sectors with the highest differences are social services, but the low number of observation at the group level (Nace 3 digits) do not allow to obtain robust indicators for the comparisons. Indicators are affected by the presence of higher values in revenues in the Sector studies compared with SMEs.

For free-lances and professional activities there is a problem relating to the application of the accounting cash method compared to the general accounting accrual method to determine the financial year. This issue makes it difficult to compare the differences between Sector studies and SMEs it depends on different interpretations of the survey questionnaires which does not present an element of systematic errors, but may generate measurement errors.

In manufacturing sectors there are no systematic differences for the following variables: Revenue, Purchases of goods and services and Value added. In some domain there are differences between SMEs and Sector studies due to anomalous cases that affected the indicators used for comparison. In all cases, however, is predominant the share of enterprises with percentage differences of  $\pm 10\%$ .

In the construction activities differences in Value added do not show a regularity neither by number of persons employed or by Nace class. There might be a different interpretation of the classification of items in the value of production (income and work in progress) between SMEs and Sector studies, but this issue do not significantly affect the comparison of revenues.

## 7. Decomposition of differences in the final estimations

At the end of the quality analysis, some variables have been directly calculated from the administrative sources and assumed as auxiliary variables  $X$ , for which differences among sources are not imputed to a different definition. They are the following:

- Turnover;
- Proxy of Value of production (including purchase of goods for resale, changes in raw materials and goods for resale);
- Proxy of Intermediate costs (including purchase of goods for resale and excluding changes in raw materials);
- Purchases of goods and services;
- Value added (difference between the proxy of the Value of production and the proxy of Intermediate costs);
- Personnel costs;
- Gross Operating Surplus (GOS).

For these variables the following five quality criteria have been satisfied:

1. total difference below 5%;
2. KS (Kolmogorov-Smirnoff indicator) not significant ;
3. Share of enterprises and the value of the variable at least 70% for the percentage difference within 5%;
4. Good consistency and no significant distortion of these indicators at the level of NACE 2-digit;
5. Good consistency and no significant distortion of these indicators at the level of the class of employees.

The other variables from administrative sources have to be considered as a proxy  $X^*$  to be used for the estimation of the  $Y$  variables coherent with SBS and SEC regulations at domain level.

The model  $Y = f(X^*)$  needed survey data to correct discrepancies in administrative data due to sectorial regulations, no perfect alignment of the definitions and linkage errors. In these case the use of auxiliary information will improve the predictive power of estimators and will open the way to two alternative research developments: mass imputation (Chipperfield et al. 2012) or aggregate combined estimators (Pfeffermann, 2003).

The process of integration of the multi-source database and the reconstruction for all the population of the main economic variables allowed us to evaluate two component in the differences in survey estimates compared with administrative-based estimates: (1) the first one is the replacement (source effect) of the survey with administrative data and (2) the second one is the sample weights (sample effect) compared to census data. The exercise has been made on the variable value added and is coherent with the previous analysis conducted within the integration of the sample of SMEs (Casciano et al., 2012).

The final dataset used in this evaluation integrated missing records or incorrect data (approximately 6%) with a cold deck imputation procedure of the per capita value of the median by stratum (Ateco three-digit group, size class and region). The imputed value is then equal to per capita median of the stratum multiplied the number of employees of the unit with missing data. The resulting values preserved the distributions of each variable in the strata. The coherence at unit level is guaranteed by the fact that the value added is recalculated after the single imputation of the value of production and of the intermediate costs.

The estimate of the Value Added based on Administrative data is the following:

$$Y_{admin} = \sum_{i=1}^N y_i^a$$

The estimate of the Value Added based on Survey data is the following:

$$Y_{sme} = \sum_{i=1}^n y_i w_i$$

The difference is decomposed in two parts, adding and subtracting the survey estimate based on administrative data ( $\sum_{i=1}^n y_i^a w_i$ ):

$$Y_{admin} - Y_{sme} = \left( \sum_{i=1}^N y_i^a - \sum_{i=1}^n y_i^a w_i \right) + \left( \sum_{i=1}^n y_i^a w_i - \sum_{i=1}^n y_i w_i \right)$$

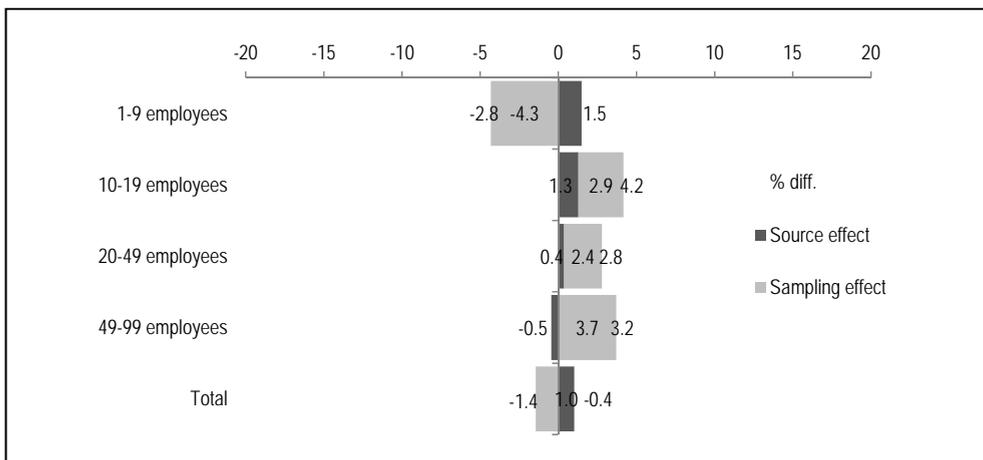
The first part represents the sampling effect in which the weights  $w_i$  are substituted with all the observations of the Business Register with complete information for the estimated variable (the missing values, about 5%, are imputed through a median imputation by stratum):

$$sampling\ effect = \sum_{i=1}^N y_i^a - \sum_{i=1}^n y_i^a w_i$$

The second part represented the source effect (with the same weight  $w_i$ ):

$$source\ effect = \sum_{i=1}^n y_i^a w_i - \sum_{i=1}^n y_i w_i$$

**Figure 4. Value added estimate in survey ( $Y_{sme}$ ), in administrative data ( $Y_{admin}$ ) and decomposition of the total difference by size, Year 2010**



The decomposition of the differences between census estimation from administrative data ( $Y_{admin}$ ) and survey calibrated estimation ( $Y_{sme}$ ), for the variable Value added (Figure 4), showed a prevalence of the so called "sampling effect" (sum of all observations with respect to the weighted sum), related to sampling error, on the "source effect" related to measurement error. The first is equal to -1.4 percentage points and the second to +1.0 percentage points, and both contribute to a difference between the estimated administrative and survey data from SMEs -0.4%. In fact the effect of sample weights have a negative impact (-4.3 percentage points) on the class 1-9 employees and is positive for the upper classes. The source effect is almost always positive: administrative-based estimate is higher than 1% and decreases with increasing company size. It went from +1.5% for micro enterprises to +0.4% for medium enterprises with 20-49 employees and became negative for those over 49 employees. At the sectorial level there is a prevalence of the sample effect in service activities and this effect is almost always opposite to the effect of replacing the data from administrative sources and is greater in sectors with a high concentration of micro-enterprises.

## Conclusions

The quality analysis has 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.

The decomposition of the differences 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 small 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 result of MNAR in the initial sample<sup>24</sup>) compared to the reference population.

---

<sup>24</sup> Missing Not at Random (MNAR): missing observations related to a specific subset of the population. It has been verified that the estimated variable depend on the response rate (F. Oropallo 2011).

## References

- Casciano M.C., A. Cirianni, V. De Giorgi, T. Di Francescantonio, A. Mazzilli, O. Luzi, F. Oropallo, M. Rinaldi, E. Santi, G. Seri, G. Siesto. 2011. *Utilizzo delle fonti amministrative nella rilevazione sulle piccole e medie imprese e sull'esercizio di arti e professioni*. Istat Working Papers n.7/2011.
- Casciano M.C., V. De Giorgi, F. Oropallo, G. Siesto. 2012. "Estimation of Structural Business Statistics for Small Firms by Using Administrative Data". *Rivista di statistica ufficiale* n.2-3/2012.
- Chipperfield J., J. Chessman, R. Lim. 2012. "Combining Household Surveys Using Mass Imputation to Estimate Population Totals". *Australian & New Zealand Journal of Statistics*. Vol. 54, Issue 2, June 2012: 223-238.
- Commission Regulation (EC) No 97/2009 of 2 February 2009 implementing Regulation (EC) No 295/2008 of the European Parliament and of the Council concerning structural business statistics, as regards the use of the flexible module.
- Commission Regulation (EC) No 250/2009 of 11 March 2009 implementing Regulation (EC) No 295/2008 of the European Parliament and of the Council as regards the definitions of characteristics, the technical format for the transmission of data, the double reporting requirements for Nace Rev.1.1 and Nace Rev.2 and derogations for structural business statistics.
- Commission Regulation (EC) No 251/2009 of 11 March 2009 implementing and amending Regulation (EC) No 295/2008 of the European Parliament and of the Council as regards the series of data to be produced for structural business statistics and the adaptations necessary after the revision of the statistical classification of products by activity (CPA).
- Commission Regulation (EU) No 275/2010 of 30 March 2010 implementing Regulation (EC) No 295/2008 of the European Parliament and of the Council, as regards the criteria for the evaluation of the quality of structural business statistics.
- Commission Regulation (EU) No 549/2013 of 21 May 2013 on the European system of national and regional accounts in the European Union.
- Istat. 2012. *Struttura e competitività del sistema delle imprese industriali e dei servizi*. Statistica Report del 29 ottobre 2012.
- Istat. 2014. *I nuovi conti nazionali in Sec 2010*. Statistica Report del 6 ottobre 2014.
- Istat. 2015. *Economia non osservata nei conti nazionali*. Statistica Report del 4 dicembre 2015.
- Luzi O., G. Seri, V. De Giorgi, G. Siesto. 2013. "Estimating Business Statistics by integrating administrative and survey data: an experimental study on small and medium enterprises". *Rivista di statistica ufficiale*, n. 2-3/2013.
- OECD. 2002. "Measuring the Non-Observed Economy - A Handbook" (OECD, 2002).
- Oropallo F. 2011. *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) – Contributi Istat – n. 7/2011*.
- Pfeffermann D. 2013. "New important Developments in Small Area Estimation". *Statistical Science of Institute of Mathematical Statistics*. Vol. 28, No. 1: 40-68.

Regulation (EC) No 295/2008 of the European Parliament and of the Council of 11 March 2008 concerning structural business statistics (recast).

Yung, W., P. Lys. 2008. *Use of Administrative Data in Business Surveys - The Way Forward* - Statistics Canada - IAOS Conference on Reshaping Official Statistics - Shanghai, 14-16 October 2008.

## Annex A

Table A.1 - Number of Asia enterprises and persons employed of the SMEs reference population by single administrative source. Year 2010.

SOURCE	Enterprises		Persons employed (no.)	Enterprises (%)	Persons employed (%)
	Total (no.)	with employees (no.)			
Financial Statements	718,538	462,650	4,940,295	16.2	39.3
SS - F	2,834,584	1,172,856	8,513,828	63.8	67.7
SS - G	661,340	116,483	930,808	14.9	7.4
PF-RG	1,347,005	331,649	2,043,266	30.3	16.2
PF-RE	592,131	88,097	742,431	13.3	5.9
SP-RG	402,471	154,826	965,249	9.1	7.7
SP-RE	39,367	18,416	143,231	0.9	1.1
SC-RS	793,130	472,712	n.d.	17.8	n.d.
PF-CM	393,048	1,113	394,580	8.8	3.1
PF-RF	183,003	106,972	641,296	4.1	5.1
SP-RF	326,217	205,838	1,584,309	7.3	12.6
SC-RF	753,135	482,232	5,070,668	16.9	40.3
<i>at least one admin source</i>	4,277,197	1,466,425	12,463,874	96.2	99.1
<i>no admin source</i>	166,686	64,685	112,582	3.8	0.9
Oros	1,516,169	1,512,825	9,119,873	34.1	72.5
SME survey	37,656	23,502	412,466	0.8	3.3
<b>SMEs population (from Asia)</b>	<b>4,443,883</b>	<b>1,531,110</b>	<b>12,576,456</b>	<b>100.0</b>	<b>100.0</b>

**Table A.2 - Number of Asia enterprise and persons employed of the SMEs reference population by overlapped administrative source. Year 2010.**

SOURCES	<i>Enterprises (no.)</i>	<i>VAT Turnover (th.)</i>	<i>Persons employed (no.)</i>	<i>Employees (no.)</i>	<i>Enterprises (%)</i>	<i>VAT Turnover (%)</i>	<i>Persons employed (%)</i>	<i>Employees (%)</i>
<b>All enterprises</b>								
<i>Financial statements</i>	718,538	1,130,303,208	4,940,295.21	4,102,499.97	16.2	64.8	39.4	58.5
SS-F	2,275,422	398,015,809	5,353,790.07	2,194,976.91	51.2	22.9	42.6	31.3
SS-G	655,566	55,185,613	920,566.62	215,064.18	14.8	3.2	7.3	3.1
PF-RG	115,605	5,685,656	151,522.47	34,211.36	2.6	0.3	1.2	0.5
PF-RE	36,731	1,391,467	37,869.66	1,412.24	0.8	0.1	0.3	0.0
SP-RG	37,299	2,851,284	79,319.23	21,201.25	0.8	0.2	0.6	0.3
SP-RE	2,045	1,368,307	7,156.51	2,916.74	0.0	0.1	0.1	0.0
PF-CM	392,979	6,832,844	394,507.19	531.57	8.8	0.4	3.1	0.0
PF-RF	7,512	3,820,944	24,284.31	14,953.19	0.2	0.2	0.2	0.2
SP-RF	17,803	21,420,833	103,302.27	72,185.60	0.4	1.2	0.8	1.0
SC-RF	17,697	78,057,218	143,394.60	123,449.28	0.4	4.5	1.1	1.8
<i>no source</i>	166,686	36,478,938	420,447.52	233,630.14	3.8	2.1	3.3	3.3
<b>TOTAL</b>	<b>4,443,883</b>	<b>1,741,412,122</b>	<b>12,576,455.66</b>	<b>7,017,032.43</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
<b>Enterprises with employees</b>								
<i>Financial statements</i>	462,650	1,042,144,197	4,660,389.09	4,102,499.97	30.2	71.1	50.8	58.5
SS-F	820,538	284,973,622	3,480,048.87	2,194,976.91	53.6	19.5	38.0	31.3
SS-G	114,762	25,147,913	353,972.75	215,064.18	7.5	1.7	3.9	3.1
PFRG	26,179	2,204,135	61,520.34	34,211.36	1.7	0.2	0.7	0.5
PFRE	1,171	143,560	2,608.83	1,412.24	0.1	0.0	0.0	0.0
SPRG	13,496	1,488,630	42,759.33	21,201.25	0.9	0.1	0.5	0.3
SPRE	722	1,215,400	4,452.95	2,916.74	0.0	0.1	0.0	0.0
PFCM	1,108	23,207	1,673.27	531.57	0.1	0.0	0.0	0.0
PFRF	3,798	3,226,327	19,968.05	14,953.19	0.2	0.2	0.2	0.2
SPRF	10,518	20,053,044	91,578.83	72,185.60	0.7	1.4	1.0	1.0
SCRF	11,483	71,413,642	136,706.16	123,449.28	0.7	4.9	1.5	1.8
<i>no source</i>	64,685	11,234,629	310,649.78	233,630.14	4.2	0.8	3.4	3.3
<b>TOTALE</b>	<b>1,531,110</b>	<b>1,463,268,306</b>	<b>9,166,328.25</b>	<b>7,017,032.43</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

## Annex B

**Table B.1 - Main indicators of the comparison between SME survey and minimum taxpayers (CM) variables (1,406 enterprises)**

VARIABLES	KS	Median difference	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	0.57	0.0	0.0	0.0	75.7	85.2	87.8	74.4	85.1	88.1
Value added at factor cost	0.45	0.0	-0.4	1.8	37.3	53.9	59.7	34.7	53.1	60.2

**Table B.2 - Main indicators of the comparison between SME survey and sole proprietorships (Modello Unico PF, RE form) variables (1,835 enterprises)**

VARIABLES	KS	Median difference	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	0.30	0.0	0.0	0.0	69.9	82.6	86.3	66.0	86.2	89.6
Other revenues and income	13.27	0.0	0.0	0.0	82.3	82.6	82.9	14.0	14.3	16.1
Purchases of goods	18,16	-26,7	-68,2	66,2	1.9	2.5	3.1	0.0	1.0	1.3
Purchases of services	4.97	1.2	-20.0	93.4	0.7	3.1	6.5	0.0	3.2	6.9
Purchases of goods and services	4.09	5.6	-5.0	64.4	7.8	16.8	22.9	2.1	28.3	36.0
Use of third party assets	16.66	0.0	0.0	0.0	62.2	64.6	66.2	26.8	41.7	48.9
Personnel costs	14.61	0.0	0.0	0.0	89.9	92.8	93.5	59.7	79.4	83.8
Value added at factor cost	0.38	0.0	-0.8	3.2	20.2	51.7	64.4	11.5	53.3	69.7
Gross operating surplus	0.50	0.0	-1-0	3.1	19.0	50.6	62.9	11.3	53.0	67.8

**Table B.3 - Main indicators of the comparison between the SME survey and sole proprietorships (Modello Unico PF, RG form) variables (6,018 enterprises)**

VARIABLES	KS	Median difference	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	0.66	0.0	0.0	0.0	75.6	89.8	91.9	71.0	91.2	93.3
Other revenues and income	18.51	0.0	0.0	0.0	80.0	80.3	80.5	6.9	8.1	9.5
Purchases of goods	9.16	0.0	-2.8	0.4	35.0	51.0	57.6	30.0	76.5	82.3
Purchases of services	12.60	49.3	1.3	176.5	4.2	7.6	11.8	2.8	7.8	14.1
Purchases of goods and services	5.0	10.6	0.9	43.7	4.8	18.3	29.0	1.8	41.6	58.1
Use of third party assets	20.8	0.0	-100	0	64.5	65.2	65.3	10.4	13.2	14.6
Personnel costs	31.2	0.0	0.0	0.0	68.7	73.0	75.1	45.0	64.8	73.8
Value added at factor cost	0.84	0.0	-6.0	1.4	14.7	42.7	56.1	9.1	44.1	58.9
Gross operating surplus	1.95	-0.2	-8.3	0.0	18.1	46.0	57.7	12.5	44.8	58.2

**Table B.4 - Main indicators of the comparison between SME survey and partnerships (Modello Unico SP, RE form) variables (410 enterprises)**

VARIABLES	KS	Median difference	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	0.17	0.0	0.0	0.0	66.1	85.6	90.2	58.6	93.8	95.2
Other revenues and income	6.95	0.0	0.0	0.0	10.3	79.7	80.5	10.3	79.7	80.5
Purchases of goods	6.84	-67.8	-94.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Purchases of services	2.73	29.6	-0.1	111.7	1.0	5.9	14.1	0.0	5.6	11.8
Purchases of goods and services	1.39	8.2	-1.9	42.7	2.9	14.1	27.3	0.2	11.4	20.6
Use of third party assets	3.92	0.0	0.0	11.2	41.0	47.6	51.5	32.0	39.4	42.5
Personnel costs	6.24	0.0	0.0	0.0	67.1	75.9	79.3	47.1	76.8	81.2
Value added at factor cost	0.42	0.2	-0.8	4.9	9.8	46.3	62.4	1.5	60.4	72.9
Gross operating surplus	0.45	0.1	-0.9	4.5	10.2	46.8	62.2	1.8	61.6	70.5

**Table B.5 - Main indicators of the comparison between SME survey and partnerships (Modello Unico SP, RG form) variables (2,780 enterprises)**

Variables	KS	Median difference	First quartile	Third quartile	% units			% value		
					±0%	±2%	±5%	±0%	±2%	±5%
Revenues from sales and services	1.24	0.0	0.0	0.0	68.6	85.7	87.7	63.3	90.7	92.7
Other revenues and income	14.53	0.0	-18.6	0.0	70.1	70.7	71.2	5.6	8.1	10.1
Purchases of goods	9.09	0.0	-3.0	0.2	36.7	52.4	58.5	29.9	79.2	85.3
Purchases of services	8.92	52.1	3.9	182.5	2.4	6.4	10.0	1.9	9.5	16.8
Purchases of goods and services	3.80	14.8	1.6	56.4	2.5	13.1	22.4	0.7	34.9	58.9
Use of third party assets	14.73	0.0	-100.0	0.0	53.3	53.8	54.1	6.5	7.4	8.5
Personnel costs	20.48	0.0	0.0	0.0	59.1	63.8	66.9	40.8	56.0	66.2
Value added at factor cost	0.75	-0.1	-8.5	3.1	8.9	34.3	48.2	5.9	34.5	49.9
Gross operating surplus	1.80	-0.98	-13.7	0.0	12.0	38.5	50.6	9.1	38.3	52.1