The new statistical register "Frame SBS": overview and perspectives[1]

Orietta Luzi,[2] Roberto Monducci[3]

Abstract

The paper summarizes the main features of the new statistical register on economic accounts of Italian small and medium enterprises (SMEs) developed at Istat in 2013-2014. The register, which is called "Frame SBS", allows to annually estimate the main variables of the economic accounts based on the massive use of microdata from integrated administrative sources. The sampling survey on SMEs is used as complementary source of information for estimating those variables which cannot be directly obtained from administrative sources. For methodological and technological details, the paper refers to the other papers published in this Volume of the Statistical Review. The paper highlights the role of the register in the area of business statistics, focusing on its potentials in terms of integrability with other sources for complex economic analyses.

Keywords: Structural Business Statistics, Administrative data, Data integration, Economic analysis.

Sommario

Questo lavoro riassume le principali caratteristiche del processo di produzione del registro statistico sui conti economici delle piccole e medie imprese (PMI) realizzato presso l'Istat nel periodo 2013-2014. Il registro, chiamato "Frame SBS", permette di stimare annualmente le principali variabili di conto economico delle imprese sfruttando massivamente microdati ottenuti da più fonti amministrative. L'indagine campionaria sulle PMI è usata come fonte complementare per la stima delle voci che non possono essere direttamente ottenute dalle fonti. Per i dettagli metodologici e tecnologici, il lavoro rimanda agli altri lavori pubblicati in questo numero della Rivista. Il lavoro evidenzia il ruolo del registro nell'area delle statistiche economiche sulle imprese, concentrandosi sulle sue potenzialità in termini di integrabilità con altre fonti per analisi economiche complesse.

Parole chiave: Statistiche Strutturali sulle Imprese, Dati Amministrativi, Integrazione, Analisi Economica

---

[1] The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat.
[2] Istat, e-mail: luzi@istat.it.
[3] Istat, e-mail: monducci@istat.it.

1. Introduction

The primary use of administrative data (*admin data* hereafter) for statistical estimation purposes in the business area is proven to produce significant benefits in terms of data quality and costs, and to greatly widen the possibilities of statistical and economic analyses. Concerning the latter, very detailed information across different dimensions (spatial-demographic and longitudinal) becomes usually available, and more reliable analyses can be performed; on the other hand, since admin data are commonly stable (over the time and in terms of information contents) they are also very suited for detailed longitudinal studies. Concerning data quality and costs, the increased completeness of statistical databases obtained on the basis of admin sources commonly produces a number of benefits as a result of the availability of census-like information on (sub-sets of) economic variables. This information also represents useful auxiliary information which can be used to increase the efficiency of samples (e.g. by focusing them on specific uncovered sub-populations or variables). In addition, in multi-source statistical databases relationships between phenomena can be estimated more reliably.

From a methodological point of view, when admin data are used as primary source of information for estimation purposes, possibly in combination with other sources, new and challenging issues need to be handled (see among others Wallgren *et al.* 2007, and Li-Chun Zhang 2012, for an overview and a discussion). Actually, since admin data are gathered for other purposes than statistical ones, additional data analysis and data processing are needed to ensure their statistical usability, such as harmonizing concepts and definitions with respect to the target statistical units and variables, matching classifications, data editing/data validation, data modelling and estimation, etc. In addition, as usually any admin source does entirely cover the variables/population of interest, information from different sources (admin archives and/or direct survey data) needs to be integrated together. In this case, additional methodological issues need to be handled in order to ensure data consistency, i.e. coherent data at micro and aggregate level. At micro-data level, micro-integration methods have been introduced, consisting in a set of approaches aiming at "*improving the data quality in combined sources by searching and correcting for the errors on unit level*" (Bakker 2010). Different micro-integration approaches can be adopted depending on the specific nature of the errors (see Li-Chun Zhang 2012, and Di Zio *et al.* 2014, for additional discussion of these issues). Moreover, as not all the target information is usually covered by the integrated sources, non-responses (either item or total) may result. Not covered information can be treated by either mass imputation (Pannekoek 2011), consisting in the prediction of microdata based on suitable statistical models ensuring consistent estimates at any level of detail, or by using appropriate estimation strategies, which exploit as much as possible the available auxiliary information (see among others Kloek *et al.* 2013).

In this paper we focus the attention on the integrated use of admin data for estimating structural business statistics (*SBS* hereafter). In this context, Costanzo (2011) provides an overview of existing practices in the European Countries, highlighting that, despite the availability of relevant, high quality and suitable admin data in most Countries, their potential is not yet fully exploited by the large majority of Member States. Portugal is the only European Country which has adopted an SBS estimation strategy where surveys have been completely replaced by admin sources directly provided by the businesses data owners

(Chumbau *et al.* 2010). Only France (Chami 2010) and Nordic Countries like Sweden, Denmark, the Netherlands and Norway (Wallgren *et al.* 2007) have implemented integrated systems for estimating SBS which are based on the primary use of admin sources, possibly complemented by sampling surveys to investigate either specific sub-populations or peculiar variables.

In Italy, in the SBS area a number of admin sources nowadays provide high quality information on businesses at high level of detail (Monducci 2010; Luzi *et al.* 2013). This fact has allowed the Italian National Statistical Institute (Istat) to gradually revise its estimation strategy in this context, moving from a production model essentially based on direct survey data complemented by admin information, to a new model where admin sources are extensively used, complemented by direct survey data. In 2013, Istat has actually developed a new statistical register for the annual production of profit-and-loss accounts of small and medium enterprises (SMEs hereafter). This register (called *Frame SBS*) massively uses firm-level admin and fiscal data as primary sources of information to estimate key SBS, while sample data on SMEs are used for estimating those items which are not available in the admin archives. Based on the *Frame SBS*, starting from the 2011 reference year, estimates of key SBS are available at an extremely refined level of detail, as they are obtained from a complete micro-data set. As expected, the design and implementation of the register has required an initial high investment in terms of methodological, technological and operational innovation, nevertheless it has determined substantial gains in terms of accuracy (as estimates of the key SBS are free of sampling errors) and consistency of estimates over time and among statistical domains (including National Accounts). The detailed and comprehensive information which is at the moment available in the *Frame SBS*, possibly integrated with other sources of information, represents a key factor in order to better analyze the characteristics and behavior of the Italian economic system (Monducci 2015).

This paper provides and overview of the main aspects which have been handled for the development of the *Frame SBS*, and illustrates the potentials benefits related to its usage in the Istat system of business statistics. Specific methodological and technological aspects are treated more in depth in the other papers included in this volume. The paper is structured as follows. Section 2 outlines the *Frame SBS*, with an overview of the main methodological and operational innovations which have been implemented for its development. In Section 3 the potentials of the new register in terms of further integration and economic analysis are discussed. Main conclusions and future work are reported in Section 4.

2. The statistical register Frame SBS

In Italy, traditionally SBS are estimated based on direct surveys. Concerning SMEs, a sample survey annually investigates about 105,000 SMEs (enterprises with less than 100 persons employed) in the industrial, construction, trade and non-financial services sectors (about 4.3 million of units as target population). In this context, admin archives are essentially used as complementary sources of information to compensate for nonresponse (see Curatolo *et al.* 2016 for a detailed description of the survey). The main issues in this

context relate to the high burden on enterprises which determines exceedingly low response rates and high sampling errors on parameters' estimates (variables' totals).

The increased stability, timeliness and quality of external admin sources on enterprises profit-and-loss accounts, such as the *Financial Statements (FS)*, the *Sector Studies Survey (SS)*, the *Tax returns forms* (UNICO) and the *Social Security data* (*SSD*) (for more details, see Curatolo *et al.* 2016 in this volume) opened the floor to a complete revision of the Istat approach to the SBS estimation on SMEs: in the new estimation strategy, admin data cover the *core* SBS information and the sample survey is used to investigate variables which are not available in the admin sources. Actually, by combining the admin sources, about 95% of the SMEs' target population is covered each year.

As the considered sources are partially overlapping (some of them provide information on the same variables on common SMEs sub-populations) a quality assessment process on each candidate data source has been initially performed, aiming at selecting the "best source" for each sub-population and variable. The evaluation process (for more details, see Curatolo *et al.* 2016 in this volume) has been based on a set of quality criteria such as *relevance*, *coverage* (in terms of target population units), *completeness* (in terms of covered information on target statistics), *accuracy, timeliness*, *integration* (extent to which the data source is capable of undergoing integration or of being integrated) (Istat 2015). The quality assessment process included a "harmonization" step aiming at reconciling the admin and the statistical definitions as described by the SBS regulation. Based on this process, "priorities" have been assigned to each source: for each source only some variables for some sub-populations of businesses have been considered reliable enough, and a specific priority among the considered sources has been established.

In Figure 1 the coverage and the priorities among sources are graphically represented: for each source p ($p=$ *FS, SS, UNICO/IRAP*), the non-dotted areas correspond to the covered SME sub-populations, identified based on the Business Register[4]. The variables Y_j^p (j=1,…,k) represent the SBS variables covered by the p considered sources at microdata level. The SSD archive contains information on employment costs for all the SMEs with at least 1 employee[5]. The variables Y_j^{SME} relate to the SBS information collected by the annual sample survey on SMEs.

Based on this framework, the annual production process of the statistical register has been implemented. It consists of three macro-phases:

1. *Sources' acquisition and standardization;*
2. *Sources' integration;*
3. *Estimation.*

2.1 Sources' acquisition and standardization

In this macro-phase (for more details, see Altarocca *et al.* 2016 in this volume) of the production process, the admin sources are acquired from the Istat centralized Business

---

[4] Containing structural information on enterprises such as: Economic Activity (Ateco), Number of Employees (Nem), Turnover (Turn).
[5] Personnel Costs (PC), Wages and Salaries (WS), Worked Hours (WH), Social Contributions (SC).

Registers sector, which is in charge of performing first data treatments and of uniquely identifying statistical units in each source. The sub-set of items required for the SBS estimation purposes is drawn from each source. Selected data are then subject to a structured set of activities aiming mainly at harmonizing variables and deriving SBS items, verifying that information relating to the same unit is unique, identify and eliminate possible inconsistent information reported within each unit, eliminate unusable information (for more details, see Sanzo *et al*. 2016 in this volume).

Figure 1 – The Frame SBS information framework

Units	ID Ateco N_{Em} Turn	N_{Em} PC WS WH SC	Y_1^{FS} Y_2^{FS} .... Y_k^{FS}	Y_1^{SS} Y_2^{SS} .... Y_k^{SS}	Y_1^{UNICO} Y_2^{UNICO} ....Y_k^{UNICO}	Y_1^{SME} Y_2^{SME} .... Y_p^{SME}
1 2 . . . . N (4.4 mil)	Business Register	Social Security Data (SSD)	Financial Statements (FS) (~16% of SMEs)	Sector Studies Survey (SS) (~80% of SMEs)	Tax Returns Data (UNICO, IRAP) (~97% of SMEs)	SME Survey (repeated)
			Not covered (~4%)			

2.2 Sources' integration

At this stage of the production process (for more details, see Altarocca *et al.* 2016 in this volume), the sources' data are combined together, and a data validation process on the integrated data is performed. The aim is to ensure the consistency of related information coming from different sources, and to identify possible measurement errors in the data. The data validation process includes the harmonization of information on employment and labour cost coming from admin and fiscal data with respect to the corresponding one coming from the *Social Security Data* (for more details, see Arnaldi *et al.* 2016 in this volume). Outliers and influential data are selected on the distributions of economic parameters like per-capita labor cost and (labour) productivity, in order to identify possible measurement errors in variables.

At the end of these processes, two sub-groups of items are identified: the first group (referred to as *main economic aggregates*) is represented by the key SBS variables[6] which are extensively covered at microdata level by the integrated sources; the second sub-group relates to the *components of the main economic aggregates,* which are characterized by inadequate coverage rates and/or quality in the admin sources.

2.3 Estimation

According to the above variables' classification, a *mixed* estimation procedure has been adopted.

For the *main economic aggregates*, a predictive approach based on *mass imputation* has naturally allowed to build a complete micro-data file: in this phase, non- available information is predicted on the basis of admin data using a combination of different imputation techniques, which are applied to separate groups of variables taking into account their distributional characteristics and their relations with other variables (for more details, see Di Zio *et al.* 2016 in this volume).

For the *components of the main economic variables*, domain estimates at the detail level required by the SBS Regulation are obtained based on a design based/model assisted approach consisting in the use of a *projection estimator* (for more details, see Righi *et al.* 2016 in this volume) which exploits the SME survey data while ensuring consistency with respect to the main economic aggregates taken as auxiliary information.

It has to be underlined that the *Frame SBS* statistical production process exploits innovative IT solutions based on the development of a new data warehouse of integrated SBS information which is well-suited for supporting the management of modules in generic workflows (for more details, see Sanzo *et al.* 2016 in this volume).

The *Frame SBS,* when combined with the data from the annual survey on profit-and-loss accounts of Large Enterprises (enterprises with more than 100 employees) currently represents the reference framework for the convergence and consistency of many surveys on specific economic topics. Moreover, its integrated use in combination with data from other statistical registers (referring to both structural and short-term trends), has opened new perspectives for Istat data users, as illustrated in the next section.
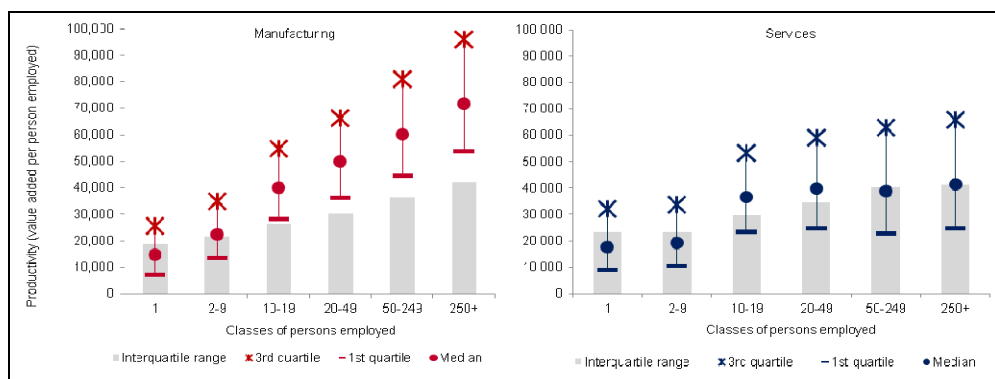
3. Potentials of the statistical register Frame SBS for economic analysis

The availability, on an annual basis, of main profit-and-loss accounts data on all companies active in Italy allows to carry out insightful analyses on both business structure and dynamics. As for the former, it is possible to assess the degree of heterogeneity within the business system, identifying the better- and worse- performing segments (e.g. sectors, clusters, etc.).

---

[6] Income from Sales and Services (Turnover), Changes in stocks of finished and semi-finished products, Changes in contract work in progress, Changes in internal work capitalized under fixed assets, Other income and earnings (neither inancial, nor extraordinary), Purchases of goods, Purchases of services, Use of third party assets, Changes in stocks of raw materials and for resale, Other operating charges, Personnel Costs.

In order to better illustrate the information potential of the *Frame SBS* register, Figure 2 reports some statistics about the distribution of the labour productivity by firms' size classes in 2013, in the manufacturing and services sectors. Besides confirming the well-known positive correlation between firm's size and productivity, the data show the heterogeneity within all size classes, revealing for instance that with the exception of the micro enterprises segment, in any other size class the most productive firms (i.e. the ones belonging to the fourth quartile of the productive distribution) perform better than the median firm of the next higher size class.

Figure 2 – Value added per person employed, by size classes – Year 2013 (euros)

Source: Istat

With regard to the dynamic analysis, the statistical register *Frame SBS* allows to longitudinally evaluate the performance of single production units, pointing out for example the firm- and sector-level developments underlying the aggregate dynamics. The latter element is particularly important for an assessment of the resilience and vulnerability of the Italian business system, as the *Frame SBS* makes it possible to monitor on an annual basis the relative competitive position of all the Italian firms within their own sector or across the entire business system, in terms of profitability, productivity and other economic performance indicators.

The register makes it possible to evaluate whether (and how) the Italian productive system that is coming out from the crisis differ from the one that entered it, for example in terms of number and size of the units, employment, and (labour) productivity. In 2010-2013 about 21% of the "persistent" firms increased the number of persons employed. From a sector perspective, the share of firms with a net job creation is higher in manufacturing (30%) than in the service sector (19,7%). These changes have partially modified the structure of Italian firms by size. In the same period, over 50% of firms increased their value added, and 15% showed a simultaneous increase in terms of value added and employment. On the other side, 43% of firms have experienced a fall both in value added and employment.

Finally, it has to be underlined that the statistical register *Frame SBS* provides a "structure information cornerstone" for further integrations with other firm-level statistical registers, referring both to structural and short-term economic events. This feature allows to identify the developments underlying some important recent trends, also taking account, in

a multidimensional way, the structural features and the strategic choices used by firms to cope with those trends.

4. Conclusions

The statistical register *Frame SBS* can be considered an advanced example in Istat of statistical production based on the direct use of multi-source, administrative data. The adoption of a mixed estimation strategy, exploiting as much as possible the available information, and the use of innovative methodological approaches for data validation, data prediction and estimation, ensure high levels of quality for the final outputs.

The benefits associated to the use of the *Frame SBS* mainly relate to the increased accuracy of cross-sectional estimates of the main SBS aggregates, and better coherence over time of the SBS estimates. At present, much more information is available not only for external data users, but also for Istat internal users, since the *Frame SBS* represents a source of auxiliary information for statistical production processes in other economic domains.

In the long term, further methodological developments are expected to produce additional benefits. In particular, a reduction of survey costs and statistical burden, associated to a further increase of data quality, will result from: the extension of the direct use of administrative data to other SBS variables and sub-populations (e.g. enterprises with more than 100 persons employed); the revision of the design of the overall SBS estimation strategy (with particular reference to the sample design for small enterprises); the adoption of innovative estimation approaches (like Small Area Estimation models, see Luzi *et al.* 2015) for specific, complex business accounts variables. A further area of development will concern the problem of identifying appropriate indicators for measuring the quality of the *Frame SBS* outputs, since they are obtained based on the direct use of multi-source administrative data. In this context, the results of European projects like the Essnet AdminData (2013) and the Essnet BLUE-ETS (2012) could be profitably exploited.

From an analytical point of view, the *Frame SBS* represents a powerful source of micro level business information for economic analysis and to support to policy advice. Its peculiar characteristics are: fully consistency with official figures, harmonization of all variables across different data sources, and totally scalable data from the micro to the meso up to the macro level of economic analysis. In addition, given its register-based nature, the *Frame SBS* represents the natural hub of an open information system that can be integrated with other variables coming from short terms survey data, dedicated qualitative surveys and policy oriented indicators. According to the growing need for micro-integrated registers designed for economic analysis, a short-term implementation plan of *Frame SBS* for economic analysis has been established, with the development of a set of further indicators aimed at assessing the competitiveness and the growth potential of Italian firms according to three relevant dimensions of enterprise's activity: employment and wages, engagement in foreign trade, business location. The output of this project will be an open system accessible through a dedicated research lab inside ISTAT where all relevant stakeholders and independent researchers will actively participate to expand and deepen our knowledge about the resilience features and evolutionary patterns of the Italian economy.

Riferimenti bibliografici

Arnaldi S., C. Baldi, R. Filippello, L. Mastrantonio, S. Pacini, P. Sassaroli e F. Tartamella. 2016. The labour cost variables in the building of the Frame. *Rivista di Statistica Ufficiale Istat,* n. 1/2016.

Altarocca F., D. Bellisai, A. Laureti Palma e R. Sanzo 2016. New experiences in the production of business statistics: the construction of the 'Frame' and the SBS-datawarehouse. *Rivista di Statistica Ufficiale Istat,* n. 1/2016.

Bakker B. F. M. 2010. Micro-Integration: State of the art. In *Report WP1: State-of-the-art on Statistical Methodologies for Data Integration*, ESSNET on Data Integration, http://www.cros-portal.eu/content/wp1-state-art.

Chami S. 2010. Reengineering French structural business statistics - an extended use of administrative data. *European Conference on Quality in Official Statistics* (Q2010), Helsinki.

Chumbau A., H.J. Pereira e S. Rodrigues S. 2010. *Simplified Business Information (IES): Impact of Admin Data in the production of Business Statistics*. Presented at the Admin Data ESSnet Seminar "Using administrative data in the production of business statistics. Member states experiences", Rome, March.

Costanzo L. 2011. *An Overview of the Use of Administrative Data for Business Statistics in Europe*. Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin (Session CPS035).

Curatolo S., V. De Giorgi, F. Oropallo, A. Puggioni e G. Siesto. 2016. Quality analysis and harmonization issues in the context of the SBS frame. *Rivista di Statistica Ufficiale Istat,* n. 1/2016.

Di Zio M., u. Guarnera e R. Varriale. 2016. The estimation of the main variables of the economic account of small and medium enterprises based on administrative sources. *Rivista di Statistica Ufficiale Istat,* n. 1/2016.

Essnet AdminData. 2013. *WP6 - Quality Indicators when using Administrative Data in Statistical Outputs, Deliverable 6.5 / 2011:Final list of quality indicators and associated guidance*, available from http://essnet.admindata.eu/WorkPackage?objectId=4257.

Essnet BLUE-ETS. 2012. *Deliverable 4.2: Report on methods preferred for the quality indicators of administrative data sources*. http://www.blue-ets.istat.it/index.php?id=7.

Kloek W. e S. Vâju. 2013. The use of administrative data in integrated statistics. *NTTS - Conferences on New Techniques and Technologies for Statistics*. Brussels, 5-7 March.

Istat. 2015. *Linee guida per la qualità dei processi statistici di fonte amministrativa*. http://www.istat.it/it/files/2010/09/LineeGuida_v.1.0_Luglio_2015.pdf.

Luzi O., F. Solari e R. Monducci. 2015. Small area estimation for business statistics: new perspectives at Istat. Invited paper. *ITACOSM 2015 - 4th ITAlian Conference on Survey Methodology*. Rome, 24-26 June.

Luzi O., U. Guarnera e P. Righi. 2014. The new multiple-source system for Italian Structural Business Statistics based on administrative and survey data. *European Conference on Quality in Official Statistics* (Q2014). Vienna.

Luzi O. e M. Di Zio. 2014. Editing administrative data. In *Memobust Handbook on Methodology of Modern Business Statistics.* http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics.

Luzi O., M. Di Zio, F. Oropallo, A. Puggioni e R. Sanzo. 2013. Integrating administrative and survey data in the new Italian system for SBS: quality issues, *3rd European Establishment Statistics Workshop (EESW 2013)*, 9-11 September. Nuremberg, Germany.

Monducci R. 2015. Measuring economic, social and environmental resilience. *Joint IEA/ISI Strategic Forum 2015 and High-Level Expert Group on the Measurement of Economic Performance and Social Progress Workshop.* http://www.oecd.org/statistics/measuring-economic-social-progress/IEA-ISI%20Strategic%20Forum%20and%20Resilience%20Workshop%20agenda.pdf. Rome, 25-26 November.

Monducci R. 2010. Statistiche ufficiali e analisi della competitività del sistema delle imprese: aspetti concettuali, problemi di misurazione, strategie di miglioramento della qualità. *Atti della X Conferenza nazionale di statistica*, Roma, dicembre.

Pannekoek J. 2011. Models and algorithms for micro-integration. In *Report on WP2: Methodological developments. Essnet on Data Integration.* http://www.cros-portal.eu/content/wp2-development-methods.

Rao J. N. K. 2003. *Small Area Estimation*. New York: John Wiley and Sons.

Righi P. 2016. Estimation procedure and inference for component totals of the economic aggregates in the new Italian Business frame. *Rivista di Statistica Ufficiale Istat,* n. 1/2016.

Wallgren A. e B. Wallgren. 2007. *Register-based Statistics: Administrative Data for Statistical Purposes.* John Wiley & Sons.

Zhang L.-C. 2012. Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*. 66; 41-63.