INTEGRATION BETWEEN DATA FROM REGISTER AND SAMPLE SURVEYS: ENTERPRISES CLASSIFIED BY USE OF ICT AND ECONOMIC INDICATORS

Alessandra Nurra, Giovanni Seri, Valeria Tomeo

Abstract. Business investments in Information and Communication Technologies (ICT) that impact production processes represent an important lever for enhancing business productivity. The data presented in this work offer new indicators and classifications by integrating the phenomenon of digitization with elements of economic performance. The Italian National Institute of Statistics (Istat) annually measures the digital transformation of enterprises with at least 10 persons employed through a sample survey that is harmonized at the European level and focuses on the use of ICT and e-commerce. The survey primarily adopts a qualitative approach, and to explore the economic characteristics of companies on the basis of their level of ICT adoption it is essential to integrate productivity and profitability indicators. Particular attention was given to the possibility of integrating the ICT data with the statistical information system, known as FRAME SBS (Structural Business Statistics), which is used for estimating structural economic variables related to business accounts on the entire population of enterprises, as defined by the SBS Regulation. The calibration estimator methodology has been chosen as the most suitable approach for integrating the data from the FRAME SBS with the ICT sample survey, aiming at analysing indicators based on data stemming from both sources. Specifically, the analysis includes the presentation of structural, productivity and competitiveness measures examined in relation to the indicators of the digital intensity of enterprises with at least 10 persons employed, such as the use of personal computers and the internet, web presence, and online sales.

1. ICT and competitiveness

Investments in Information and Communication Technologies (ICT) that impact production processes play a crucial role in driving business productivity growth. The new Regulation on European business statistics emphasizes the significance of measuring the digital economy and the use of ICT, recognizing their influence on competitiveness, growth, and the need to promote European strategies and policies related to the completion of the digital single market. Both the European

Commission and Parliament highlight the importance of monitoring the digital progress of Member States through the Digital Economy and Society Index (DESI). Additionally, there is a focus on measuring EU trajectories against targets defined by the Digital Decade Policy Programme 2030.

The Italian National Institute of Statistics (Istat) annually measures the digital transformation of enterprises with at least 10 persons employed through a sample survey that is harmonized at the European level. This survey focuses on the utilization of ICT and aims to investigate the extent of adoption of various emerging technologies, considered by policymakers as enabling behaviours and virtuous processes that can enhance the competitiveness of businesses. The survey primarily adopts a qualitative approach, and it is essential to integrate indicators of productivity and profitability. This integration allows for an examination of the economic characteristics of companies based on their level of ICT adoption. Furthermore, it enables the creation of an integrated database suitable for cross-sectional and panel analyses, particularly for larger enterprises included in the survey (all enterprises with at least 250 persons employed are included in the ICT survey).

In literature, there are numerous studies related to the impact that the adoption of information and communication technologies (ICT) can have on innovation processes, production mechanisms (such as robotics or additive manufacturing), organizational functions, as well as firm and overall system performance. However, this positive relationship between ICT usage and firm productivity has not always been consistently demonstrated, particularly in advanced countries. This has led to the development of two distinguishable schools of thought known as "techno-optimists" and "techno-pessimists" (Andrews *et al.*, 2016; Cette *et al.*, 2016). As a result, some studies argue, for various reasons, that Italy belongs to the countries where the use of ICT struggles to become a lever for improving the entire economic system. This could be attributed to factors such as the prevalence of small enterprises that hinder the diffusion of new technologies (Accetturo *et al.*, 2013), inefficiencies in management selection (Pellegrino and Zingales, 2017), limited investment in human capital (Bugamelli and Pagano, 2004), or the low effectiveness of innovation support policies (Bronzini and Piselli, 2016).

In recent years, Istat has also conducted several analyses on the competitiveness of Italian companies in relation to their levels of digitization and capital endowments (both human and physical) at both national and regional levels. These analyses have identified, on one hand, the presence of virtuous digitization behaviors, and on the other hand, the need for adequate physical and, more importantly, human capital inputs to effectively transform ICT into growth opportunities. Special attention was dedicated to exploring the potential integration of data between the statistical information system, known as FRAME SBS (Structural Business Statistics), which estimates structural economic variables based on business accounts, and sample

surveys such as the one focused on the use of ICT (Istat, 2020). The FRAME SBS variables are derived from administrative data or statistical estimation methods. They constitute a comprehensive archive that encompasses the entire population of enterprises as defined by the SBS Regulation.

2. Methodology

The aim of this work is to integrate the information gathered by a sampling (ICT) and exhaustive (FRAME SBS) sources by combining qualitative indicators and economic variables through a method qualifying the results in terms of comparability and consistency according to Istat statistical standards.

Different approaches have been considered: macro methods such as balancing and iterative proportional fitting, and micro methods including statistical matching with or without weights, consistent repeated weighting, and calibration estimators (Seri *et al.*, 2016). The calibration estimates methodology was applied as the most suitable approach for integrating data from the FRAME SBS with the ICT survey. It utilizes the interaction between an exhaustive register and sample data to produce economic indicators. This is the same methodology used in the production process of the ICT sample survey. However, the initially proposed set of indicators does not directly or indirectly replicate published estimates. Instead, it utilizes the information derived from the combination of the two data sources in a manner that ensures substantial or complete consistency with both sources. Considering the significance of the Digital Intensity Index (EDII) as an indicator used by Eurostat, based on the use of 12 digital activities and now being employed for the Digital Decade, we have also incorporated this indicator into our analysis.

Several differences in applying method are listed as follows:

- considering t as reference year of the ICT survey, the statistical archives used for estimations consist of FRAME SBS refers to the previous year (t-1), while the ICT sample survey of year t utilizes ASIA referred year (t-2). The given population is identified based on updated information from the reference year (t-1);
- a small percentage of enterprises interviewed by the survey are no longer part of the considered population, mostly because they have fewer than 10 persons employed. These units have been excluded from the sample, and the data set used for indicator estimations is based on matching observations between the ICT survey units and the units in the FRAME SBS. Anyway, more than 90% of the ICT respondents is eligible for the analysis;
- the estimation domains are redefined by matching the survey's required domains (with small details) to align with the study's objectives (tables of

indicators with specific characteristics) and the quality of the results obtained. Specifically, the territorial level has been reduced, excluding regional information;

- in terms of the calibration model used for weights in the ICT survey (totals for variables such as the number of enterprises and number of employees by NACE and geographical level), the utilization of FRAME SBS was crucial to consider known totals related to Value Added, Turnover, and Gross Operating Margin.

The methodological framework utilized is structurally the same as that of the ICT survey. Therefore, to assess the accuracy and precision of the produced estimates, the same criteria accompanying the currently published estimates can be adopted. The ReGenesees software (Zardetto, 2015), which implements the methods commonly used in Istat for economic surveys, was employed for the analysis. Furthermore, since the adopted strategy generates microdata files with a weighting system that represents the entire population (similar to the survey) it was possible to reproduce the estimates of the ICT indicators (which is not the objective of this work to replicate). These estimates are entirely consistent with the published ones, providing reassurance about the consistency of the results (the consistency with FRAME SBS is guaranteed by design in the new estimation domains). However, it is important to reiterate that the objective of this work was to define a series of indicator tables that combine information from FRAME SBS and the ICT survey. Therefore, it is advisable to exclusively use the dataset of elementary units (and their corresponding weights representing the entire population) that enabled the achievement of this objective.

2.1. ICT indicators

The ICT indicators have been constructed in such a way as to divide the population of enterprises into exhaustive classes according to different types of ICT adoption. As previously mentioned, the proposed set of indicators does not replicate, either directly or indirectly, published estimates, but it leverages the informative interaction between the two sources in substantial or complete coherence with both. Additionally, for enterprises with at least 10 persons employed, the structural characteristics are presented (export propensity, group membership), as well as productivity and efficiency (labor productivity) and competitiveness (labor cost per employee) indicators in relation to indicators measuring the extent of ICT utilization.

The ICT indicators are as follows:

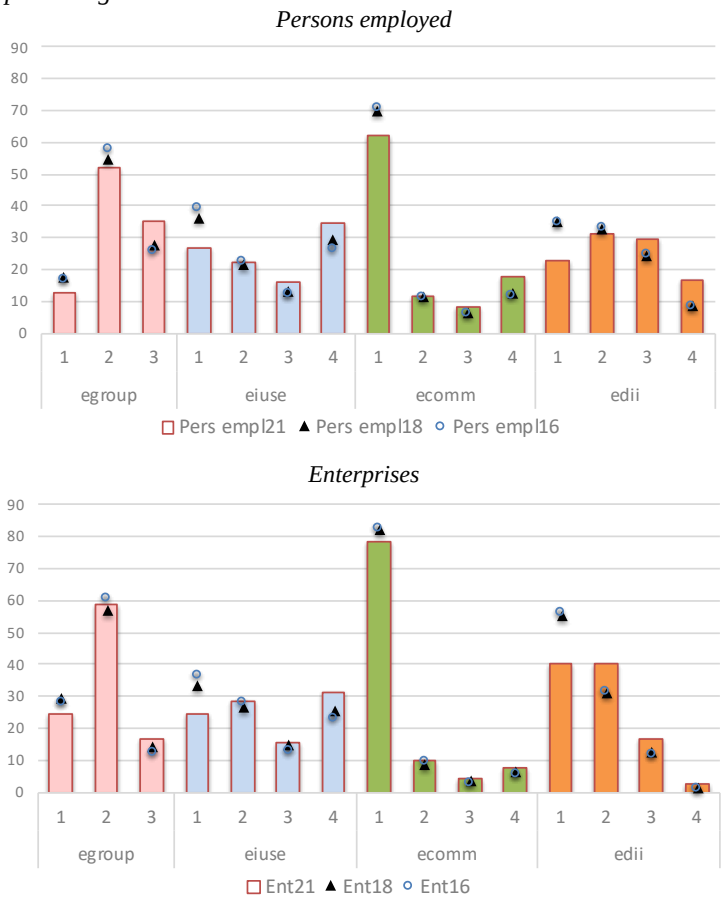
Table 1 – *Indicators that measure the extent of ICT adoption.*

ICT Indicators	Values	Description
EGROUP		Enterprises classified by level of Internet usage (PC online, web, e-sales)
	1	Enterprises without/with PC and Internet but not having a website and online sales
	2	Enterprises connected to the Internet and having a website but not making online sales
	3	Enterprises connected to the Internet and making online sales
EIUSE		Enterprises classified by penetration rate of Internet usage among persons employed
	1	Enterprises without PC and Internet + those using PCs and less than 25% persons employed are connected to the Internet
	2	Enterprises with % of persons employed using Internet between 25% and less than 50% of the total
	3	Enterprises with % of persons employed using Internet between 50% and less than 75% of the total
	4	Enterprises with at least 75% of persons employed using the Internet
ECOMM		Enterprises classified by selling online or not
	1	Enterprises not having website or having it but without shopping cart functionality and not selling online
	2	Enterprises having website with shopping cart functionality but not selling online in previous year
	3	Enterprises selling online in previous year without shopping cart functionality (indirect e-commerce proxy)
	4	Enterprises selling online in previous year with shopping cart functionality
EDII		Digital intensity index used by Eurostat based on use of 12 digital activities
	1	Very low intensity – if enterprises use 0-3 digital activities
	2	Low intensity – if enterprises use 4-6 digital activities
	3	High intensity – if enterprises use 7-9 digital activities
	4	Very high intensity – if enterprises use 10-12 digital activities

3. Key findings

Data reveals the increasing evolution of digital usage (edii2 and edii3, eiuse4) and the decreasing of enterprise that have low use of Internet (egroup1, eiuse1, ecomm1, edii1). Confirming ICT annual results, data reveal that only few enterprises are engaged in e-commerce (ecomm2, ecomm3, ecomm4) and in very high level of ICT activities adoption (edii4). (Figure 1)

Figure 1 – *Persons employed and enterprises by ICT indicators. Year 2016, 2018 and 2021, percentage values.*



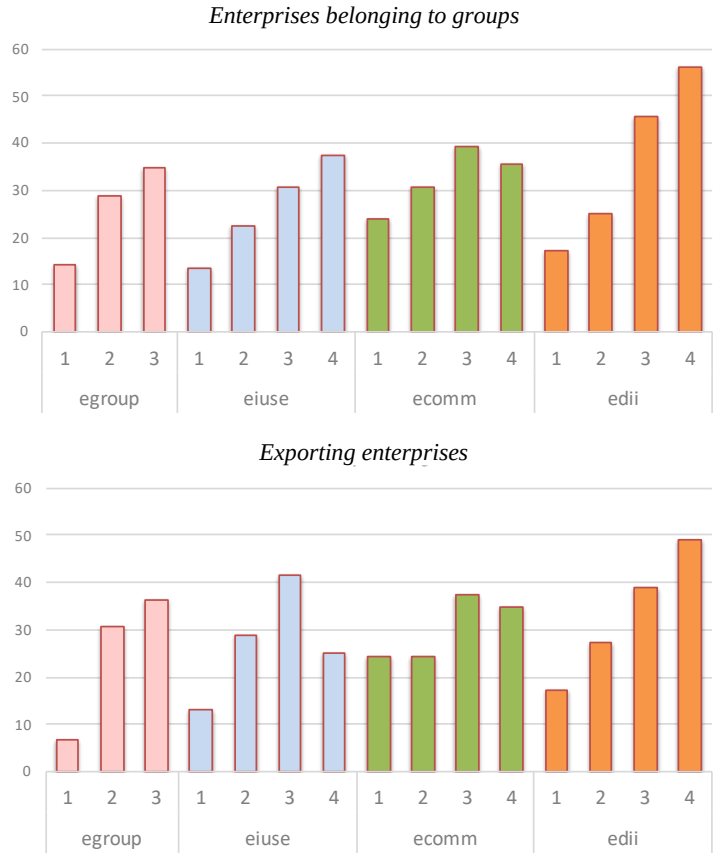
Authors' elaborations on Istat data

The share of companies with at least 10 persons employed belonging to groups increases as it grows digitization indicators or if they use third-party online sales channels beyond their own (ecomm3).

The same occurs for exporters, except for the last level of the eiuse indicator, where there is a prevalence of companies with high technological content and low propensity to export, such as software production and IT consulting.

A similar situation is observed for the level 4 of ecomm indicator, in which companies are concentrated in the accommodation services sector. (Figure 2)

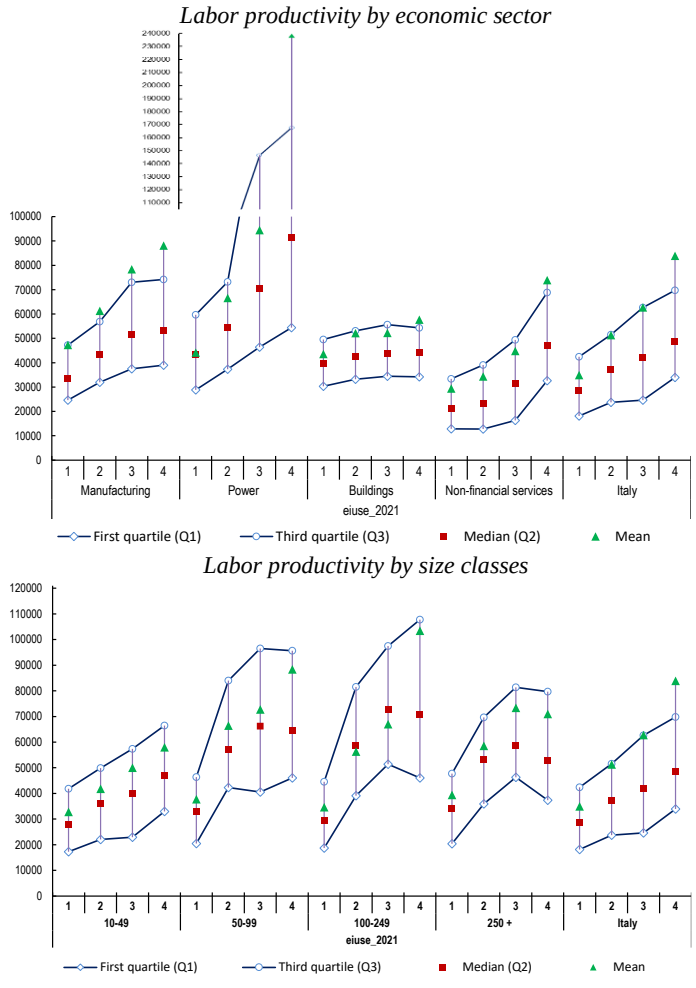
Figure 2 – *ICT indicators by structural characteristics (belonging to groups and exporting enterprises). Year 2021.*



Authors' elaborations on Istat data

There is a positive correlation between the adoption of technologies (eiuse) and the labor productivity measured in terms of value added per persons employed, especially in manufacturing and energy companies and, in particular, for those already in a position of high productivity compared to the others (third quartile). This positive correlation is evident also in terms of classes of persons employed and is very similar in all the years analysed. (Figure 3)

Figure 3 – *Labor productivity and levels of Internet usage (eiuse) by economic sector and by size classes of persons employed. Year 2021.*



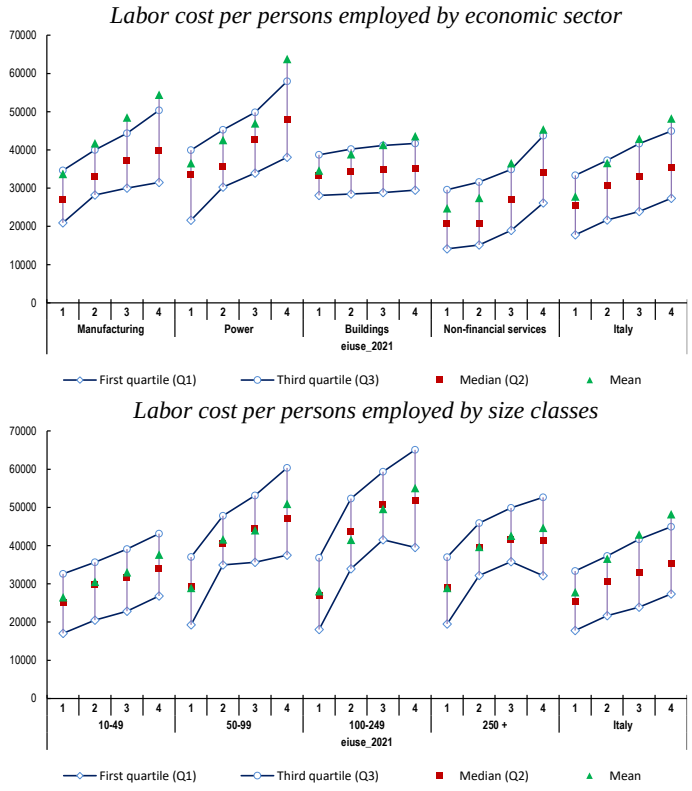
Authors' elaborations on Istat data

The interquartile gap seems to increase mostly as the technology indicator rises, highlighting a greater dispersion around the average of the indicator and therefore a greater differentiation between those who get the most benefits and those who fall behind.

In general, the percentage growth of the economic indicators considered at the level of persons employed class or macro-sector shows larger increases, especially during the initial transition from low or no ICT usage to the subsequent level. However, the relationship between ICT and economic indicators becomes less evident in subsequent jumps and, at times, even negative.

The positive correlation is evident also between the adoption of technologies (eiuse) and the labor cost per persons employed in terms of economic sector and size classes of persons employed. (Figure 4)

Figure 4 – *Labor cost per persons employed and level of Internet usage (eiuse) by economic sector and by size classes of persons employed. Year 2021.*



Authors' elaborations on Istat data

When analysing the data regarding the distribution of companies across quartiles of labor productivity, a general trend emerges: there is an increase in the proportion of companies belonging to higher levels of the chosen ICT indicator as the quartile of economic performance grows (from the first to the second and third quartile). The share of larger-sized companies is greater than that of smaller-sized companies in the higher quartiles of productivity, particularly at higher levels of ICT adoption (Figure 5). Smaller-sized companies show productivity improvements at lower levels of ICT adoption compared to what is required for larger-sized companies.

Figure 5 – *EDII indicator levels by quartiles of labor productivity and size classes of persons employed. Years 2018 and 2021.*

p.e. classes	quartile classes	edii_2018 1	2	3	4	edii_2021 1	2	3	4
10-49	<=Q1	74,1	19,5	5,8	0,6	46,6	40,4	11,0	2,1
10-49	Q1 - Q2	55,6	31,3	12,3	0,9	48,4	40,1	10,8	0,7
10-49	Q2 - Q3	54,9	31,8	12,4	1,0	42,4	40,0	15,3	2,3
10-49	>Q3	45,4	39,2	14,0	1,5	32,7	42,6	22,6	2,2
50-99	<=Q1	64,0	27,6	6,6	1,8	46,0	33,4	16,6	4,0
50-99	Q1 - Q2	43,0	37,3	17,5	2,2	26,3	40,1	29,6	4,1
50-99	Q2 - Q3	37,4	39,7	19,3	3,6	11,9	37,7	39,9	10,4
50-99	>Q3	27,2	41,3	26,7	4,8	7,0	35,8	49,7	7,5
100-249	<=Q1	61,7	26,4	11,0	0,8	25,5	36,5	26,9	11,0
100-249	Q1 - Q2	30,9	43,4	23,6	2,0	15,1	30,8	40,0	14,1
100-249	Q2 - Q3	18,6	48,0	29,3	4,1	5,6	23,3	48,6	22,5
100-249	>Q3	16,8	39,6	34,9	8,7	2,2	16,2	49,6	32,0
250 +	<=Q1	43,9	38,6	15,5	2,1	48,1	35,1	13,2	3,6
250 +	Q1 - Q2	18,1	37,3	36,5	8,1	28,4	46,9	20,3	4,5
250 +	Q2 - Q3	12,1	33,7	43,7	10,5	19,3	42,2	36,1	2,4
250 +	>Q3	4,9	30,2	47,8	17,1	12,8	42,8	37,0	7,3
Total	**<=Q1**	**73,0**	**20,4**	**5,9**	**0,7**	**46,7**	**39,7**	**11,2**	**2,3**
Total	**Q1 - Q2**	**55,0**	**31,6**	**12,5**	**0,9**	**47,3**	**39,7**	**12,0**	**1,0**
Total	**Q2 - Q3**	**52,0**	**33,1**	**13,5**	**1,4**	**38,9**	**40,7**	**17,7**	**2,7**
Total	**>Q3**	**40,3**	**39,6**	**17,6**	**2,5**	**27,5**	**41,4**	**26,9**	**4,1**

Authors' elaborations on Istat data

4. Results and future work

The integration of microdata between the ICT survey and the SBS Frame has allowed for expanding the range of available indicators by leveraging the combined information from the two sources.

During the calibration process, the regional level of estimation detail was lost, but it was possible to maintain the other estimation domains by fixing the known totals to other economic variables.

The methodology presented here for the ICT survey has also been applied to other surveys, such as the Community Innovation Survey (CIS), and has provided equally interesting results.

References

ACCETTURO A., BASSANETTI A., BUGAMELLI M., FAIELLA I., FINALDI RUSSO P., FRANCO D., GIACOMELLI S., OMICCIOLI M. 2013. Il sistema industriale italiano tra globalizzazione e crisi. *Questioni di Economia e Finanza, Occasional Papers 193*, Banca d'Italia.

ANDREWS D., CRISCUOLO C., GAL P. 2016. The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy. *OECD Productivity Working Papers 2016/05*, OECD Publishing, Paris.

BRONZINI R., PISELLI P. 2016. The Impact of R&D Subsidies on Firm Innovation, *Research policy*, Vol. 45, No 2, pp. 442-457.

BUGAMELLI M., PAGANO P. 2004. Barriers to Investment in ICT, *Applied Economics*, Vol. 36, No 20, pp. 2275-2286.

CETTE G., FERNALD J.G., MOJON B. 2016. The Pre-Great Recession Slowdown in Productivity, *Working Paper 2016/08*, Federal Reserve Bank of San Francisco.

ISTAT. 2020. Integration between data from register and sample surveys: enterprises classified by use of ICT and economic indicators, Experimental Statistics. Roma: Istituto Nazionale di Statistica.

PELLEGRINO B., ZINGALES L. 2017. Diagnosing the Italian Disease. *Working Paper 23964*, National Bureau of Economic Reasearch.

SERI G., ICHIM D., LUCHETTI F., COSTA S., NURRA A., MASTROSTEFANO V., SALAMONE S., PASCUCCI C., ORSINI D. 2016. Integrazione del Frame con altre Indagini e Fonti Amministrative ai fini della Produzione di Indicatori Complessi, *Istat Working Papers 17*, Istat.

ZARDETTO D. 2015. ReGenesees: An Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys, *Journal of Official Statistics*, Vol. 31, No 2, pp.177-203.

Alessandra NURRA, Istat, nurra@istat.it
Giovanni SERI, Istat, seri@istat.it
Valeria TOMEO, Istat, tomeo@istat.it