ANALYSIS OF ACCOUNTING DATA FROM THE EXPLANATORY NOTES TO THE FINANCIAL STATEMENTS FOR THE ESTIMATION OF THE ENTERPRISES INVESTMENTS[1]

Antonio Regano, Valeria Tomeo, Roberta Varriale

1. The usability of a new administrative source in Istat production processes for estimating Investments

The Italian national institute of statistics (Istat) produces estimates on the main economic variables of enterprises, both from a Structural Business Statistics (SBS) and a National Accounts (NA) perspective. The two production processes present similarities and discrepancies. Both processes use survey data and administrative data. Survey data are collected through the Business Account Survey (also known as SBS surveys), divided into the Large Enterprises (LE) census survey on all enterprises with 250 employees and over, and the Small and Medium-sized Enterprise (SME) survey on a sample of enterprises with less than 250 employees. Administrative data come from different sources and are used in different steps of the estimation process (frame definition, editing and imputation, production of estimates, etc.). Among the administrative sources, the Financial Statements of enterprises represents one of the most important, including a very rich source of information: beside filling in the accounting forms, firms report explanatory notes that include a summary of significant accounting policies and details of the reported values in their financial statement and clarifications on the economic situation of the company. Because of the different informative content, we will refer to the accounting forms of Financial Statements (*FS* hereafter) and Explanatory Notes to the Financial Statements (*Explanatory Notes* hereafter) as different administrative sources.

As far the production process of estimates on "acquisitions of fixed assets" during the year, also known as "enterprise investments" (*Investments* hereafter), both SBS and NA processes use survey data, but only NA process also uses administrative information. In particular, FS does not report information on Investments, while the Explanatory Notes do: these are used in the NA production process together with other administrative sources, but not by SBS. More specifically, SBS impute total

---

non response by donor, and, only for the large enterprises, compare the results with longitudinal data; NA use auxiliary information from administrative sources to edit influential errors on respondent data and impute total non response by applying a selective editing procedure.

The aim of the present work is to evaluate the content of the information on Investments, at the level of the legal unit, from the administrative source of the Explanatory Notes, beside the administrative sources already used in Istat processes. This work represents an exploratory and study phase to evaluate the usability of this source in the production processes of Investments of both SBS and NA. In Section 2, we describe the available administrative sources in Istat on the Investments. Section 3 compares the information on Investments from the Explanatory Notes and SBS survey, both from a structural and an empirical point of view. Section 4 describes the current NA production process of estimates on Investments by using different administrative sources. Section 5 concludes the work.

2. Availability of administrative sources on Investments

Information on Investments is present in different administrative sources: Table 1 reports the administrative sources available in Istat and their content related to the target variable Investments.

As introduced, one of the most important administrative sources in the business context is the Financial Statements of enterprises. In FS, Investments are those costs that do not exhaust their usefulness in a single administrative period, but manifest their economic benefits over a period of time that covers several financial years. These are divided into Tangible Assets, characterized by the requirement of materiality (for example machines, plants, vehicles, etc.), and Intangible Assets, which are fixed assets without physical consistency (for example patents, plant and expansion costs, goodwill, etc.). The value of the stock of fixed assets at the end of the year is reported in the balance sheet of the FS, while the details on their movement during the year (acquisitions, disposals, depreciation, write-downs, transfers, etc.) must be reported in the Explanatory Notes. Istat has access to the Explanatory Notes of corporations and limited enterprises containing information on the target variable Investments from Infocamere (in the form of XBRL[2] files and according to FS national standards) and from CRIF (for enterprises that compile FS according to the International Accounting Standard IAS or International Financial Reporting Standards IFRS). In the following, we will use the terms XBRL and CRIF to refer to Investments from the Infocamere and CRIF source, respectively.

---

[2] XBRL (eXtensible Business Reporting Language) is a global framework and a standards-based way to communicate and to exchange business information.

Table 1 – *Available administrative sources and variables related to business Investments.*

Legal form	Accounting Principles	Data provider	Variable in the administrative source	Administrative source	Consistency between available variable and Investments
Enterprises on the Basis of Liabilities	National standards	Infocamere	XBRL Investments (Splitted in Tangible and Intangible Assets)	Explanatory Notes	Total
			Delta Stock	FS	Partially
	IAS or IFRS 16-17	CRIF	CRIF Investments (Splitted in Tangible and Intangible Assets)	Explanatory Notes	Total
			Delta Stock	FS	Partially
Enterprises on the Basis of Members		Agenzia dell'Entrate	Expenses for amortizable goods (VAT)	Value Added Tax declarations	Partially

Beside the Exploratory Notes, information on Investments are available in Istat from other administrative sources, as shown in Table 1. The first one is the information on expenditure for amortizable goods (*VAT variable* hereafter) reported in Value Added Tax declarations of Agenzia delle Entrate (AA.VV., 2020). The variable VAT is likely to be a good proxy of the target variable Investments only when business investment refers mainly to: purchases of land and/or assets produced on own account and/or assets acquired when acquired through a financial lease. The second available information related to Investments is a derived variable that can be computed from FS (*DELTA_STOCK variable* hereafter), based on the assets at the end of the year minus the assets at the end of the previous year (i.e. at the beginning of the year) plus depreciation and revaluation (AA.VV., 2020). In enterprises' financial statements, acquisition of fixed assets (i.e., Investments) is one of the components that explain the difference between the value of net asset at the beginning of the accounting period and the value at the end of the accounting period. Therefore $NETSTOCK_{t,end} = NETSTOCK_{t,beg} + INVE_t + REVAL_t − SOLD_t - AMOR_t - WOFF_t + MA_t$, where $NETSTOCK_{t,end}$ and $NETSTOCK_{t,beg}$ are, respectively, net assets at the end and at the beginning of year t, $INVE_t$ is the acquisition of investments in year t, $REVAL_t$ is the revaluation of existing assets in the year t, $SOLD_t$ is the net book value (i.e., net of cumulated depreciation) of existing assets sold in the year t, $AMOR_t$ is depreciation of existing assets in the year t, $WOFF_t$ is write-off of existing assets in the year t and MA_t is the effect of mergers and acquisitions. Then from FS, we can compute the proxy of investment: $DELTA\_STOCK_t = NETSTOCK_{t,end} - NETSTOCK_{t,beg} + AMOR_t + WOFF_t$. Since $DELTA\_STOCK_t = INVE_t + REVAL_t -$

$SOLD_t$ + Ma_t, then $DELTA\_STOCK_t$ is a good proxy of investment when revaluations and selling of existing assets and mergers and acquisition are not very important.

Not all enterprises are covered by all sources of information: over 80% of the enterprises in Asia 2017 are linked to at least one of the selected administrative sources; about 63% of enterprises with less than 250 employees are linked to a single source, while 76% of enterprises with more than 250 employees are present in three sources; 19% of enterprises with less than 250 employees are not present in any source.

3. Investments in Explanatory Notes and SBS surveys

In this Section, we describe the comparison between the information on Investments from the Explanatory Notes in XBRL and SBS survey, from two different points of view: the structure of the information, and empirical results.

Istat collects the administrative source FS from Infocamere, which manages the data of the Chambers of Commerce. The data supply consists of over 900,000 FS, together with their Explanatory Notes; a first supply is in September of the year *t+1* (where year *t* is the reference for the data) and a second in March of the year *t+2*. The level of detail and the instructions for filling in the LE and SME questionnaires relating to the Investments are not completely consistent with the accounting rules of enterprises. Table 2 compares the information in the LE and SME surveys with that one in the Explanatory Notes. With regard to Tangible Assets, the main difference between the LE and SME questionnaires and the Explanatory Notes is that in the SME questionnaire, the item "Tangible assets in progress and advances" (var11) is not present. This implies that enterprises respond to the questionnaire in two ways: *i*) by not considering the value of assets under construction and advances (and therefore reporting an overall value of the acquisitions of assets different from that recorded in their financial statements); *ii*) by attributing the value of the acquisitions of fixed assets in progress to the related assets (and therefore reporting in the questionnaire an overall value of the acquisitions of fixed assets equal to that recorded in their financial statements, but with different values in terms of detail by type of asset) (Brunaccini F. *et al.*, 2016).

Table 2 – *Comparison between the detailed items of the acquisitions of Tangible and Intangible Fixed Assets (Investments) between the LE and SME surveys and the Explanatory Notes.*

	LE	SME	Explanatory Notes
		Tangible Fixed Assets	
var1	Land	Land	
var2	Dwellings	Dwellings and Other buildings (var2)	Land and Dwellings
var3	Other buildings		
var4	Machinery and equipment	Machinery and equipment	Machinery and equipment
var5	Industrial and commercial facilities		Industrial and commercial facilities
var6	ICT equipment	ICT equipment	
var7	Forniture	Forniture	
var8	Transport fixed equipment	Transport equipment	Other tangible fixed assets
var9	Valuable goods	Valuable goods	
var10	Other tangible assets		
var11	Tangible fixed assets in progress and advances		Tangible fixed assets in progress and advances
var12	Total Tangible Fixed Assets	var1+var2+var4+var6+var7+var8+var9	Total Tangible Fixed Assets
		Intangible Fixed Assets	
var13	Industrial patent rights	Artistic and literary originals, patents (var14)	Industrial patent rights and rights of use of intellectual works
var14	Rights of use of intellectual property (artistic and literary originals)		
var15	Concessions, licenses and trademarks (excluding software)		Concessions, licenses and trademarks
var16	Software	Software	
var17	Research and development		Research and development
var18	Other intangible fixed assets	Other intangible fixed assets (var18)	Start-up costs Plant expansion costs Intangible fixed assets in progress and advances Other intangible fixed assets
var19	Total Intangible Fixed Assets	var14+var16+var18	Total Intangible Fixed Assets
var20	var12+var19	Total Fixed Assets	Total Tangible Fixed Assets + Total Intangible Fixed Assets

In addition, therefore, to the mismatch between the SME data and those of the FS and the loss of homogeneity between the respondents to SME, there is also a lack of homogeneity with respect to the LE survey, in which the item is instead present. As far the information on Intangible Assets is concerned, however, the LE and SME

questionnaires have a very different structure from that typical of financial statements. In SME, information is requested only for three asset types: "Artistic and literary originals, patents" (var14), "Software" (var16) and "Other intangible fixed assets" (var18); the first two do not have a correspondence with the typical items of the Explanatory Notes, but are rather "of which", while it is required to include all the other intangible asset types in the third item "others".

The LE questionnaire is closer to the representation in the Explanatory Notes, but does not consider the items relating to "Stat-up costs", "Plant expansion costs" and "Intangible fixed assets in progress and advances". Another possible problem with the LE questionnaire is related to software investments. In correspondence with the item "Software", it is requested to include both the purchased and the self-produced one, while it is specified to exclude software with reference to the item "Concessions, licenses and trademarks" (var15). In reality, according to what is reported by the OIC (the Italian Standards Setter) accounting principles, enterprises can capitalize the software purchased as "Concessions, licenses and trademarks" or as industrial patent rights and intellectual property rights (var13 and var14), depending on the method of purchase, while software produced for own use (and capitalized) should be included in "Industrial patent rights" if protected by copyright laws and in "Other intangible fixed assets" (var18) if not protected. Therefore, specifying "excluding software" only with reference to the item "Concessions, licenses and trademarks" can be misleading and can generate inconsistency with what is reported in the Explanatory Notes.

The analysis of the empirical results in this Section uses data from 2017, to compare the information on Investments from the Explanatory Notes and the LE and SME surveys, already available for the reference year. For the surveys, only the respondent enterprises were selected for a total of 37,729 units, 34,790 for SME and 2,939 for LE. By integrating this dataset with that of the Explanatory Notes, which consists of 941,184 observations after the pre-treatment operations, the reference universe dropped to 21,136 enterprises, 18,945 of SME and 2,191 of LE.
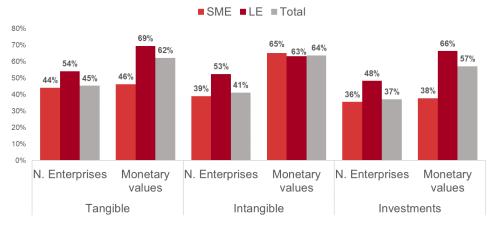
Figure 1 reports the comparison between the data extracted from the Explanatory Notes and those collected by the SME and LE surveys: the bars represent the percentage incidence in terms of number of enterprises and their monetary values of the units with a difference of ± 5% between SBS and Explanatory Notes sources. Figure 1 highlights:

   - for the total investments, coherence between the data of the administrative source and those of the surveys at the microdata level which stands at 37% in terms of the number of enterprises on the entire sample;
   - for the total investments, 57% of the monetary values indicated by the enterprises fall within the considered range, a share which rises to 66% for LE and stops at 38% for SME;

- considering the SME survey, the monetary values for intangible fixed assets perform better results than intangible (65% vs 46%).

On the other side, the critical issues concern both the component relating to tangible fixed assets for the SME survey, due to the lack of the detail item relating to fixed assets in progress and advances, and the intangible component for both surveys, due to a different reporting scheme compared to as adopted in the Explanatory Notes; and the presence of possible measurement errors and zero values in one of the two sources.

Figure 1 – *Enterprises and monetary values in the range of percentage difference of ± 5% between sources, by type of investment and survey (percentage incidence). Year 2017.*

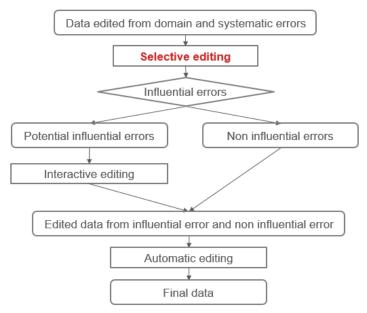*Source: elaboration on SBS and Explanatory Notes data. Authors' elaboration.*

4. NA production process for estimating Investments

The NA current procedure for estimating Investments is different for the subset of units belonging to the SME and LE population. Since the aim of the present work is to evaluate the content of the information on Investments, at the level of the legal unit, from the administrative source Explanatory Notes, we focus on the LE population. Indeed, the LE is a census survey and it allows making evaluations at micro level on all the population without taking into account the sampling error.

The NA production process of Investments for the enterprises of the LE survey exploits the auxiliary information from administrative sources to correct the survey data by applying a selective editing procedure.

Figure 2 represents the NA editing and imputation (*E&I* hereafter) model flow for the estimation of Investments, based on that one proposed by the Generic Statistical Data Editing Model (GSDEM; UNECE, 2019), that is the reference framework for statistical data editing developed by UNECE. In GSDEM, the E&I process is interpreted as a set of standardized, consistently described information objects that are the inputs and outputs in the design of the overall data processing flow. In Figure 2, process steps are represented by rectangles, data states are represented by ellipses, a trivial process control is represented by an arrow and non-trivial process control is represented by a diamond.

Figure 2 – *NA editing and imputation model flow for Investments.*

Source: Authors' elaboration.

After the control for domain and systematic errors, selective editing makes it possible to localize (potential) influential errors, i.e. those observations that have the greatest influence on the estimates. The method is implemented by using the SeleMix package of R. After this step, the (potential) influential errors are evaluated and treated with a manual check. Subsequently, the phase of imputation of the total missing responses follows, by using auxiliary variables coming from the available administrative sources. The latter step is run with an automatic editing procedure. In

the NA current production process, the sources providing information on Investments are: CRIF, VAT and DELTA_STOCK variables (Section 2).

As mentioned, the aim of the present work is to evaluate the content of the information on Investments, at the level of the legal unit, from the administrative source Explanatory Notes. Therefore, the same methodology for selective editing is applied: the auxiliary administrative source XBRL is used in addition.

Before applying the selective editing model implemented in the Selemix package of R, some pre-processing step has to be done, such as the distinction between missing values and "genuine zeros" in the target variable Investments.

The method for selective editing implemented in the SeleMix package is based on a latent class model: a Gaussian model is assumed for true data and an "intermittent" error mechanism such that a proportion of data is "contaminated" by an additive Gaussian error. Details on Selemix package can be found in Guarnera and Buglielli (2020). The practical steps in the application of selective editing by using SeleMix are: analysis of data in order to choose the response variables Y and verify if auxiliary information X (covariates) is available; estimation of the contamination model parameters (estimation phase); identification of critical units corresponding to the most influential errors (prediction and selection phase); interactive editing of critical units and automatic editing of non-critical ones.

In order to avoid singularities, the contamination model is applied to all the data characterised by an observed value of the Investment greater than zero and different - up to a tolerance factor - from the values of all the covariates (VAT, DELTA_STOCK and XBRL). Specifically, the survey data where the Investment value agrees with at least one of the covariates are considered correct (not subject to the selective editing procedure) and are excluded from the estimation process (Di Zio M. et al., 2015). Another important aspect in the application of selective editing in this context, is the one concerning the different missing patterns for the covariates, in fact for different units different sets of covariates may be available. This requires the estimation of different models for different patterns, where, for each pattern, the units included in the estimation process are all those where (at least) the covariates corresponding to the current pattern are available.

In addition to the three covariates, an always observed stratification variable is also used, so that the above procedure is applied separately within each stratum. The stratification variable is the enterprise size in terms of number of employees. Precisely, two size classes are used ("250-499" and ">499").
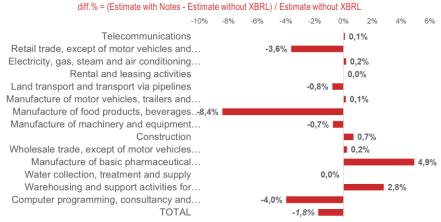
A final remark is about the treatment of missing data in the target variable. The models used for selective editing are also applied in order to impute missing data in the target variable. The estimation domains on which the impact of errors has been evaluated are 64 industries corresponding to the classification of economic activity

A*64 that is used to disseminate NA data; the threshold used for the selection of critical units has been set equal is 4% (Di Zio M. *et al.*, 2015).

4.1 Changes in the NA estimates of Investments

Figure 3 represents the percentage differences, by industry, which are most significant in terms of the incidence of the value of Investments.

Figure 3 – *Difference % between the estimate of Investments by using the XBRL source and the estimate without XBRL, by industry. Enterprises with over 250 employees. Year 2017.*

diff.% = (Estimate with Notes - Estimate without XBRL) / Estimate without XBRL

Source: elaboration on LE data. Authors' elaboration.

Results are obtained comparing two models, which use Investments from the LE survey as target variable. Model 1 uses the VAT variable as the main covariate together with DELTA STOCK and CRIF and Model 2 uses XBRL as an additional covariate. Overall, the estimate of the amount of Investments using Model 2 is 1.8% lower than the estimate with Model 1. The results reveal an overestimation of the acquisitions of fixed assets in the model using as the main covariate the VAT variable, which has some definitional differences compared to the target estimate variable. Furthermore, the estimate has a partial conceptual coherence (expenditure for amortizable goods do not totally coincide with the acquisitions of firms, they could be more consistent among small businesses) and tends to overestimate.

The estimates computed with the use of the Explanatory Notes (Model 2), together with the other administrative sources, produced overall satisfactory results for almost all domains.

5. Results and future work

The aim of this work was to analyse the information contribution of the administrative source Explanatory Notes, with respect to the Investments made by enterprises during a specific reference year. In particular, the reported analyses represent an important exploratory phase to evaluate the usability of this source in the production processes of Investments of both SBS and NA.

After mapping of the available administrative sources on Investments in Istat, we compared the data extracted from the Explanatory Notes and those collected from the LE and SME surveys on the economic accounts of the enterprises to assess the degree of matching between the sources. An empirical analysis was run on data from 2017. The results can be used to harmonize the information collected through the surveys' questionnaires with that extracted from the Explanatory Notes. Subsequently, we studied the use of the Explanatory Notes in the production process for estimating Investments by industry, adopted by the NA.

All the exploratory analyses provided satisfactory results, which need in-depth study in the future. Possible future implications may be the use of the Explanatory Notes in the editing and imputation process of SBS surveys, a methodological alignment between SBS and NA for the estimation of Investments, and, finally, the inclusion of the Investment variable in the Frame-SBS (Extended Register of the main economic variables of enterprises), representing an object of the Istat integrated system of registers.

References

ARNALDI S. 2020. Il Trattamento delle Fonti Amministrative per la Costruzione del Sistema Informativo sui Risultati Economici delle Imprese Italiane. *Working Papers 6*, Roma, Istat.

BRUNACCINI F., FORESTIERI P., IOMMI M., REGANO A. 2016. Analisi del Contenuto delle Note Integrative ai Bilanci Civilistici per la Stima delle Acquisizioni di Immobilizzazioni. *Working Papers 22*, Roma, Istat.

DI ZIO M., GUARNERA U., IOMMI M., REGANO A. 2015. Selective editing of business investments by using administrative data as auxiliary information. In *UNECE Work Session on Statistical Data Editing*, Budapest, Hungary.

GUARNERA U., BUGLIELLI M.T. 2020. *Package 'SeleMix', Selective Editing via Mixture Models.* Available on line at https://cran.r-project.org/web/packages/SeleMix/SeleMix.pdf.

UNECE 2019. *Generic Statistical Data Editing Model GSDEM (Version 2.0, June 2019).* Available on line at: https://statswiki.unece.org/display/sde/GSDEM.

SUMMARY

The Italian national institute of statistics (Istat) produces estimates on the main economic variables of enterprises, both from a Structural Business Statistics (SBS) and a National Accounts (NA) perspective. The two production processes present similarities and discrepancies. Both processes use survey data and administrative data. Survey data are collected through the Business Account Survey, divided into the Large Enterprises (LE) census survey on all enterprises with 250 employees and over, and the Small and Medium-sized Enterprise (SME) survey on a sample of enterprises with less than 250 employees. Administrative data come from different sources and are used in different steps of the estimation process. Among the administrative sources, the Financial Statements of enterprises represents one of the most important, including a very rich source of information: beside filling in the accounting forms, firms report Explanatory Notes that include a summary of significant accounting policies and details of the reported values in their financial statement and clarifications on the economic situation of the company.

As far the production process of estimates on "acquisitions of fixed assets" during the year, also known as "enterprise investments", both SBS and NA processes use survey data, but only NA process also uses administrative information.

The aim of the present work is to evaluate the content of the information on Investments, at the level of the legal unit, from the administrative source of the Explanatory Notes, beside the administrative sources already used in Istat processes. This work represents an exploratory and study phase to evaluate the usability of this source in the production processes of Investments of both SBS and NA. Possible future implications may be the inclusion of the Investment variable in the Extended Register of the main economic variables of enterprises, representing an object of the Istat integrated system of registers.

_____

Antonio REGANO, Istat, regano@istat.it
Valeria TOMEO, Istat, tomeo@istat.it
Roberta VARRIALE, Istat, varriale@istat.it