

Total Process Error framework: an application to economic statistical registers

Roberta Varriale, Fabiana Rocci, Orietta Luzi

1. Introduction

As many other National Statistical Institutes (NSIs in the following), in recent years Istat has given new impetus to the renewal of its overall strategy for the production of Official Statistics. In this strategy, the production of the required outputs in all the statistical production areas is obtained based on the combined use of both primary and secondary sources of information. Primary data are those obtained by direct surveys, while secondary data correspond to information that are made available to NSIs by external bodies, and that are used by NSIs for statistical purposes (Memobust, AA.VV. 2014). Actually, one of the fundamental principles of the new Istat production strategy is the massive and integrated use of micro data from administrative sources (hereafter AD), which are used in particular for the construction of statistical registers. Besides other methodological aspects, this deep change in the statistical production paradigm requires to adapt standards and tools for the evaluation and documentation of data quality for the final users of the registers outputs and, more generally, of the outputs of multisource processes.

In this context, the Total Process Error (TPE) framework has been recently proposed in literature for assessing the quality of multisource processes, such as the production process of statistical registers. TPE framework can be used both to support the multisource process design and to monitor an overall production process, and can provide key elements for the assessment of the quality of both the processes and their statistical outputs.

In this paper, we describe how the TPE framework can be used referring, as a case study, to the Istat Register for Public Administrations. The production process of this register is still under construction, and is characterized by a modular structure depending on the different sub-population covered by the register itself. By using the TPE, we focus on the different steps and critical “decision points” of the production process for the different modules of the register. In section 2, we describe the main elements of the TPE, in section 3 we describe its application to the Register for Public Administrations.

2. The Total Process Error framework

Total Process Error (TPE) framework has been recently proposed in literature for assessing the quality of multisource processes (Rocci *et al.*, 2022). The TPE framework represents an evolution of the Zhang’s two-phase life-cycle approach (Zhang, 2012).

The TPE includes two phases of assessment, that can be described as: Phase 1. Assessment of single data sources w.r.t. original source purposes; Phase 2. Combination/re-use/integration of data sources w.r.t. target statistical purposes, that can be further splitted in: Phase 2a. Assessment of single data sources w.r.t. target statistical purposes and Phase 2b. Assessment of the combined data sources w.r.t. target statistical purposes. For each phase, some potential errors that may arise together with specific indicators to assess them are identified.

The TPE also includes an operative tool to connect the steps of a multisource production process to the phases of the quality evaluation framework: actually, this tool consists of a cross-classification scheme describing the link between the process steps of an entire production process and the above mentioned phases of the TPE framework. The cross-classification scheme may be used both to support the design of the statistical production process and to monitor the whole process once it has been put into production. Furthermore, the scheme allows to use the TPE in a

Roberta Varriale, ISTAT, Italian National Institute of Statistics, Italy, varriale@istat.it

Fabiana Rocci, ISTAT, Italian National Institute of Statistics, Italy, rocci@istat.it

Orietta Luzi, ISTAT, Italian National Institute of Statistics, Italy, luzi@istat.it

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Roberta Varriale, Fabiana Rocci, Orietta Luzi, *Total Process Error framework: an application to economic statistical registers*, pp. 147-151, © 2021 Author(s), CC BY 4.0 International, DOI 10.36253/978-88-5518-461-8-28, in Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci (edited by), *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the on-site conference*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-461-8 (PDF), DOI 10.36253/978-88-5518-461-8

very flexible way to represent different production processes. Table 1 shows the cross-classification scheme for a multisource production process using AD composed by N steps.

Table 1. Cross-classification scheme: production process steps vs TPE phases

Process step	Phase		
	1. Assessment of single AD w.r.t. administrative purposes	2. Combination/re-use/integration of AD w.r.t. target statistical purposes	
		2a. Assessment of single AD w.r.t. target statistical purposes	2b. Assessment of the combined AD w.r.t. target statistical purposes
1			
2			
...			
N			

3. The register for public administrations, territorial bodies

The economic Register for Public Administrations (hereafter Frame PA) is the result of an Istat project started in 2019. Frame PA is a *satellite* register of the *base* Register of Public Administrations (S13 hereafter). The latter defines the Italian public administrations as a subset of the Italian business register units. The difference between *base* and *satellite* register is in the role they play in the statistical production system, given the target (sub)populations and variables they are referred to. Following Wallgren and Wallgren (2014), we can define the base registers as the ones that represent the statistical reference populations for all the statistical processes (individuals/households, economic units, etc.) and the satellite registers as those releasing additional variables usually representing specific phenomena. The information contained in the final statistical Register Frame PA will be, for each statistical unit, both structural information coming from the Register S13, and some economic variables respecting accountancy definitions.

Frame PA includes different subpopulations. Nowadays, Istat is working on the subpopulation of Local Authorities, including municipalities, unions of municipalities, provinces, mountain communities, metropolitan cities, regions and autonomous provinces.

The first step to build Frame PA for Local Authorities (hereafter Frame PALA) is to select the statistical units from the Register S13, together with some structural information (address, number of employees, etc). Subsequently, information from AD sources is extracted, integrated and treated to produce the final output, that are some economic variables according to the statistical target accountancy definitions. The main AD sources concerning the economic variables of Local Authorities are the Public Administration Database (BDAP), and the Information System on the Operations of Public bodies (SIOPE). BDAP records the accounting variables of balance sheets according to the Financial Statement Management Schemes; SIOPE is a system of digital collection of profits and payments made by treasurers and cashiers of all Public Administrations. Both BDAP and SIOPE can be queried at different times of a reference year to acquire periodic updates.

Following the subject matter experts' indications, taking into account the target population and variables of the Frame PALA, the BDAP has been defined as the primary source of information, as it provides information consistent with the statistical target accountancy definitions. This choice implies that, after drawing and integrating information from BDAP and SIOPE, missing information in BDAP need to be estimated (imputed), by using SIOPE as auxiliary variables.

Different features of BDAP source characterize the Local Authorities: information on

municipalities, unions of municipalities, provinces, mountain communities and metropolitan cities is affected by total missing values, while information on regions and autonomous provinces (22 bodies in total) usually do not suffer of this problem.

Three variables are considered in the process, both on the revenues and the expenses sides. Let Y_1^{BDAP} , Y_2^{BDAP} and Y_3^{BDAP} , with $(Y_2^{BDAP} + Y_3^{BDAP}) = Y_4^{BDAP}$, be the variables observed in BDAP and Y_4^{SIOPE} the variable observed in SIOPE corresponding to Y_4^{BDAP} . The revenues and expenses are specified in Frame PALA across 148 and 22 “items”, respectively, that are grouped in Titles. We will refer to the 148 and 22 items as the Frame PALA “theoretical scheme”.

In case of total missing values from BDAP, such as for municipalities, unions of municipalities, provinces, mountain communities and metropolitan cities, missing information in BDAP have to be fully imputed, by using SIOPE information as auxiliary variables.

Table 2 shows the coverage of BDAP at different times during 2020 and 2021 for units belonging to the base Register S13 population. The reference year for data of both Register S13 and BDAP is 2019.

Table 2 – Coverage of BDAP source with respect to the target population (Register 2013), for Local authorities type – Number of respondents. Year 2019.

Local authorities type	Total population, Register S13, 2019	Respondents, July 2020	Respondents, October 2020	Respondents, May 2021
Provinces (excluded autonomous provinces) and metropolitan cities	100	80	93	98
Municipalities	7914	6455	7521	7806
Mountain communities	151	62	71	83
Unions of municipalities	562	282	324	363
Total	8727	6879	8009	8350

The presence/absence of total missing values in BDAP, makes the design of the Frame PALA production process for the two groups of local authorities completely different. Tables 3 and 4 show how the cross-classification scheme may be used to support the design of these two production processes.

Without going into the details of the two production process steps, it is clear that the process relating to the population of municipalities, unions of municipalities, provinces, mountain communities, metropolitan cities is more complex, and comprehend both an integration and an imputation step that are not present in the production process of Frame PA for the populations of regions and autonomous provinces. This means that this process is characterised by additional critical “decision points” and potential errors that may arise. The indicators linked to these steps (and phases) will be useful to support the design of these two different production processes (Rocci *et al.*, 2022).

Table 3. Frame PA, municipalities, unions of municipalities, provinces, mountain communities, metropolitan cities: production process steps vs TPE phases.

Process step	Phase		
	1. Assessment of single AD w.r.t. administrative purposes	2. Combination/re-use/integration of AD w.r.t. target statistical purposes	
		2a. Assessment of single AD w.r.t. target statistical purposes	2b. Assessment of the combined AD w.r.t. target statistical purposes
1	Quality assessment of each candidate AD source (<i>BDAP</i> , <i>SIOPE</i>)		
2		Quality assessment of each AD source (<i>BDAP</i> , <i>SIOPE</i>) in terms of Frame PA purposes	
3			Integration of AD sources (<i>BDAP</i> , <i>SIOPE</i>), by following a “theoretical scheme”
4			Imputation of the total missing values of the variable Y_4^{BDAP}
5			Imputation of the (totally) missing values of the variables Y_1^{BDAP} , Y_2^{BDAP} and Y_3^{BDAP}
6			Computation of the output Frame PA variables as aggregation of Y_i^{BDAP} values for different items

Table 4. Frame PA, regions and autonomous provinces: production process steps vs TPE phases.

Process step	Phase		
	1. Assessment of single AD w.r.t. administrative purposes	2. Combination/re-use/integration of AD w.r.t. target statistical purposes	
		2a. Assessment of single AD w.r.t. target statistical purposes	2b. Assessment of the combined AD w.r.t. target statistical purposes
1	Quality assessment of each candidate AD source (<i>BDAP</i>)		
2		Quality assessment of <i>BDAP</i> source in terms of Frame PA purposes	
3			Transformation of the <i>BDAP</i> variables in the Frame PA output variables (computation of the output Frame PA variables as aggregation of Y_i^{BDAP} values for different items)

In the future, Frame PA will comprehend additional statistical populations, characterized by a different structure of information sources. Therefore, the production process of the output

economic variables will have different steps and critical “decision points”. TPE was a useful tool in the design phase of the Local Authorities component, it will be used in the design phase of the other components and will also be used for their monitoring once it is put into production.

References

- AA.VV. (2014). *Memobust Handbook on Methodology of Modern Business Statistics*. Available at: <https://ec.europa.eu/eurostat/cros/system/files/Memobust%20glossary%20def.pdf>.
- Rocci F., Varriale R., Luzzi O. (2022). Total process error: An approach for assessing and monitoring the quality of multisource processes. Forthcoming in *Journal of Official Statistics*, June 2022.
- Wallgren, A. and B. Wallgren. (2014). *Register based statistics: Administrative data for statistical purposes*. John Wiley & Sons, Ltd.
- Zhang, L.C. 2012. Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, **66** (1), pp. 41-63.