DE GRUYTER OPEN

Journal of Official Statistics, Vol. 34, No. 2, 2018, pp. 573–580, http://dx.doi.org/10.2478/JOS-2018-0027

Letter to the Editor

On Statistical Unit Errors in Business Statistics

Official business statistics are crucial for decision makers, but what is actually meant by a "business" has been debated in the statistical community for decades. Economic data can be based on many different unit types, such as legal or fiscal units in administrative data, and different kinds of reporting units in the case of primary data collection. The choice of the targeted business unit type is a cornerstone decision that determines the outcome of business statistics. However, despite the fundamental importance of business units in business statistics, traditionally, there has been little statistical theory and methodological guidance available regarding the prevention and treatment of errors associated with the business units. To improve the situation, we believe methodologists need to become fully aware of the nature of unit errors and their potential consequences. The aim of this letter is three-fold: (i) to summarise the core aspects of the unit error and the associated unit problem, (ii) to stimulate the discussions to clarify and improve our understanding of the system of statistical units, which is needed for the production of National Accounts and various relevant national and international business and economic statistics, and (iii) to call for a general approach to the unit problem in business statistics, including the development of necessary statistical methodology for evaluating and treating unit errors from a total survey error perspective.

1. Some Background and Recent Initiatives

Business statistics refer to populations of businesses, but the notion of "business", and its measurement, are not clear-cut. In fact, several business concepts, so-called (types of) statistical units, are used in business statistics. As the application of statistical units determines the outcome of business statistics, this subject has been discussed by the international statistical community for many years now. Several efforts to harmonise concepts and measurement have been undertaken. In the European Statistical System (ESS), this has resulted in the adoption of a regulation on statistical units (EEC 1993).

There are good reasons for wanting to apply various unit concepts in business statistics. However, there are also several types of errors associated with the definition and measurement of units. These errors, henceforth broadly called unit errors, are important. Despite the attention that the international statistical community has given to the subject of statistical units, statistical methodology has thus far been lagging in providing theory and

Acknowledgments: An earlier version of this letter was prepared as "Statement on the Unit Problem in Business Statistics" (Lorenc et al. 2017), on invitation from the Scientific Committee of the 2017 European Establishment Workshop, held in Southampton, UK, from August 30 to September 1, 2017, https://statswiki.unece.org/display/ENBES/EESW17. The authors benefited from the comments of Tihomira Dimova, Sanjiv Mahajan, Rami Peltola and Norbert Rainer in early stages of the work on the Statement. Opinions expressed in this letter are those of the authors, and have not been reviewed or endorsed by their employers.

methodological guidance on the treatment of business units in the production and use of business statistics.

The international statistical community clearly does not consider the issue of statistical units solved. Within the ESS, the Business Statistics Directors Groups (BSDG) and the Directors of Macroeconomic Statistics (DMES) have called for a fundamental discussion on statistical units in its Notice of Intention (Eurostat 2015). In 2016, this call was supported at the Fifth International Conference on Establishment Surveys (ICES V) in Geneva, Switzerland in a keynote speech by Head of National Accounts at OECD Van de Ven (2018) and also at an invited session called "The future of statistical units", in which the relationship between statistical units used in business statistics on the one hand and National Accounts on the other, was an important topic (Struijs 2016). Subsequently, the United Nations Statistics Division created the Intersecretariat Working Group on National Accounts (ISWGNA) Task Force on Statistical Units in the National Accounts, which is chaired by Peter van de Ven. This topic is also discussed at the level of National Statistical Institutes (Sturm 2015).

2. Administrative vs. Statistical Business Units

There is a distinction between administrative and statistical business units. Administrative units are created for administrative purposes outside the statistical system. For instance, legal units (LeU) are a type of administrative units that one would find in every country (although the definition varies between countries). Another example is tax units, which exist in some countries, that are created for taxation purposes and do not necessarily coincide with the legal units. Administrative business units are generally maintained by external owners and imported to the statistical system more or less frequently. They are also the starting points for creating and maintaining statistical business units.

Statistical business units are created within the statistical system for the purpose of producing statistics. Typically, intrinsic relationships between statistical units are inferred and articulated in terms of a classification, or a model of units. For example, for business statistics purposes, the current Eurostat set of statistical units (EEC 1993) comprises the unit types enterprise group (EG), enterprise (ENT), local unit (LoU), kind of activity unit (KAU), and local kind of activity unit (LKAU) (see Figure 1). Or, as another example, the Institutional Unit, which is closely related to the ENT, may be subdivided into units of homogeneous production (UHP) for the purpose of National Accounts, where the UHPs are not the same unit type as the KAUs. The different types of units are depicted with their one-to-many or many-to-many relationships in Figure 1.

3. Needs and Challenges Regarding Statistical Units

The creation of statistical units is necessary because administrative units do not exist for statistical purposes nor are they seen as being able to fully meet the needs of users of statistics. For instance, many economic theories are based on the assumption that business units possess a level of autonomy in decision-making. In contrast, LeU administrative units entail legal (or fiscal) accountability, the structure of which does not necessarily coincide with that of business decision-making.

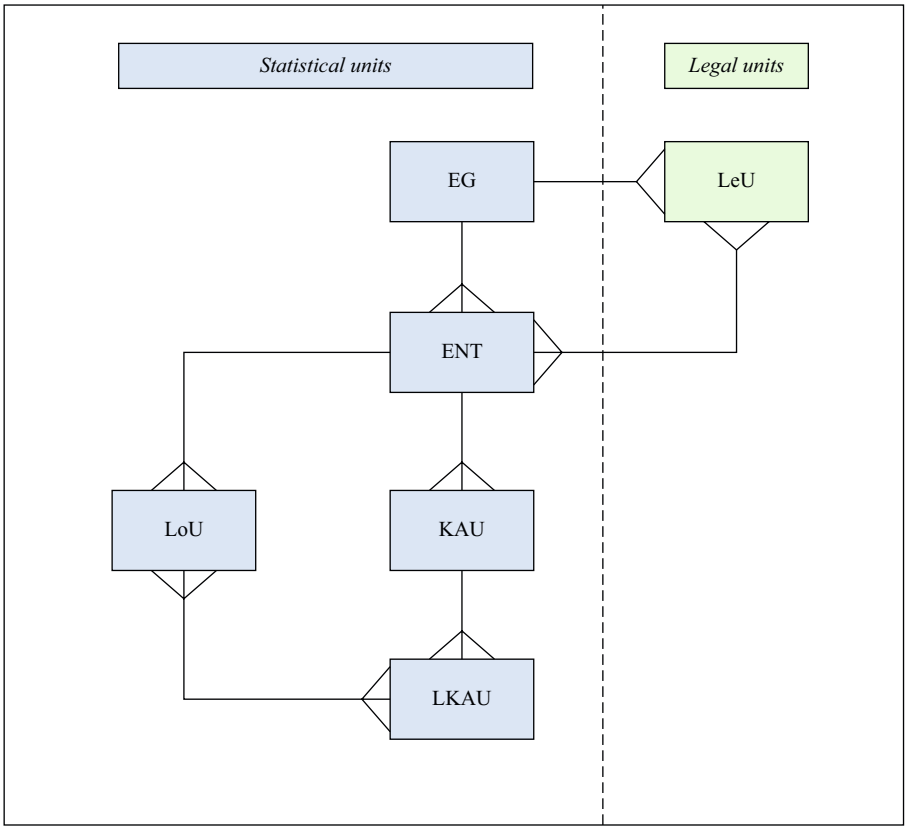

Fig. 1. A system of statistical units (ESS Task Force on Statistical Units, 2014).

Ideally, the system of statistical units should mirror, as well as possible business data availability, so as to improve data collection. For instance, the EG and ENT have been designed to capture the business management, operation and accounting structure better than the LeU. However, in practice, this is not always achieved. Some of the lower-level statistical unit types, in particular, may present challenges to businesses' own understanding of reality. For example, the business may find it difficult to extract the required data (e.g., sales, purchases, or profit) for the LKAU, in which case they may deliver data that refer to a proxy unit (e.g., LeU) instead.

In the Survey Methodology framework, which is based on sampling theory, one distinguishes between the study unit (of the target population), the sampling unit, the reporting unit (i.e., the entity within a business that is expected to actually deliver the required data), and others. The system of statistical units, which has been developed to fit business statistics and National Accounts, is based on economic theory. The survey methodology approach and the economic theory approach result in unit types that do not fully align with each other, in the sense that there does not always exist a many-to-one mapping from a type of unit under one approach to another type of unit under the other approach. Moreover, administrative units are relevant with respect to survey compliance, or the reduction of response burden and survey cost by the increasing uptake of administrative data.

4. Unit Error and Unit Problem

The choice of the business unit type is a cornerstone decision in the design and operation management for a statistical product. Recognising its importance, we are highlighting here the concept of the "unit error". Unit error refers to errors in the statistical output that are caused by discrepancies in identification, characterisation and delineation of the relevant statistical units and the relationships between them. In addition, by the term "unit problem" we refer to the challenges and obstacles to our understanding of the unit error and our efforts to deal with it. The unit problem may be related to the practices of generating the business statistical products, the design of relevant statistical processes, as well as the conceptualisation of the system of statistical units.

5. Generic Situations for Unit Errors

Unit errors can be appreciated in terms of discrepancies between "what one aims to obtain" and "what one obtains". Below are some generic situations where these discrepancies arise.

1. There is **observation error** in the obtained data, such as when a value is missing or misreported. The discrepancy is between the results based on the true data and the erroneous data. For example, the administrative record shows that an LeU is active in the economic sector "12345", whereas it is in fact active in the sector "21345". More on various kinds of observation errors is found in Section 6.
2. **Implementation error** may be the case with respect to the relevant regulation or statistical unit model (e.g., the one in Figure 1). The discrepancy is between the results from correct and incorrect implementation. For instance, the legal framework for statistical business registers may be misinterpreted, or it may not cover the extra complications in a given country (e.g., existence of tax units in addition to LeUs), and so on.
3. There may exist inconsistencies and shortcomings in the statistical unit model or relevant regulation, which we refer to as the **definition error**. The discrepancy is between the results with and without such inconsistencies. For example, the unit model depicted in Figure 1 does not include the means to guarantee that one obtains the same ENTs directly from the LeUs or indirectly via the EGs. In another example, the definition actually allows an LoU to have activity in different locations (towns).
4. There is ultimately the discrepancy between the ideally delineated units under a consistent unit model and the units that the users need or expect for their purposes. For instance, Brion et al. (2014) have documented such discrepancies between the actual business demography of small and medium-sized enterprises (SMEs) in France, which is based on the LeUs, and the user expectation of business demography based on autonomous units like the ENTs. One may refer to this as the **conceptualisation error**. As long as there is a (non-negligible) conceptual mismatch, improving the implementation of the existing relevant regulatory framework cannot overcome the unit error in statistical products. More discussions are given in the Village Bakery example below.

6. More on Observation Error

Leaving aside the implementation (including derivation rules), definition and conceptualisation errors for the moment, Van Delden (2017) documents many types of observation error. To start with, profiling is the activity to manually delineate large or complex businesses, for instance ones that operate globally. Erroneous statistical units resulting from profiling tend to have a large impact on the output.

Next, the data required by definition may be missing or simply unavailable. For example, turnover of the Value Added Tax (VAT) units may need to be transformed to that of ENTs, where the two unit types have many-to-many relationships. Or, consolidation is needed to exclude internal flows from the values reported by LoUs that all belong to the same ENT, see Haag (2018). (De-consolidation is the opposite situation.) Or, the quarterly total from an administrative source may need to be apportioned to monthly totals required by the survey.

Finally, the observed values may themselves be erroneous, which can cause various issues. For instance, the units may be identified by key variables such as "name", "postal code", and "phone number" and errors in their values (e.g., misspellings) may cause the units to be wrongly identified. For classification of a statistical unit that consists of multiple LeUs, one usually selects the main category (e.g., NACE code), which may be affected by the errors in relevant data. There may be linkage errors between an enumerated set of units and other data sets that contain information about the relations between these units. The information on the relationship between units may be mistaken, for example unit A has 70% of the shares (level of ownership) in unit B, whereas it should be 50%. Almost all the data available for the operations above may be subjected to delays in reporting or updating. (Sometimes, it is not known which time the available data refers to.) For instance, the information on the change of ownership may be delayed, or the values of identification variables may be outdated (old ID values are reported).

7. More on Conceptualisation Error: An Example of the Village Bakery

In reality, the target statistical units often need to be "created", which ultimately raises the issue of potential conceptualisation error. Let us illustrate this with a "small" example.

A bakery in a village, a family business, is organised as two LeUs, one for the operational business (producing and selling baked products) and one for the real estate (building). The reason for this arrangement is to reduce risk and have something set aside for inheritance (not subject to business liabilities, etc., and good for tax reasons). This bakery is, in fact, also a small supermarket. In any case, the statistics authority views this business as a single ENT, based on autonomy in decision-making.

The baker and his wife work daily in the bakery and shop. The wife is officially employed; the baker is the owner and manager. His sister does bookkeeping and administration for him (as well as for her husband, the baker's brother-in-law), which she is paid for, but not as an employee. Occasionally (mainly weekends and evenings), their son helps in the shop or runs errands (e.g., transport to or from the shop). He is not formally employed, as he is a secondary school pupil. In deciding what products to bake, the baker takes into consideration the production of his brother-in-law, who is a farmer; in turn, the brother-in-law takes the needs of the baker into account when deciding on next season's crops. In short, they coordinate their business activities, in addition to the engagement of the baker's sister.

Their work is done by four persons, though not all of them full-time; the location is the real estate (a section of the family home); the capital comes from running the business, plus some years ago, the baker borrowed money from a bank to modernise his production line, which he is now repaying; the raw material input (flour) comes mainly from the brother-in-law; the retail products come mainly from other local producers, plus some selected distributors; energy comes in part from the local wind farm, and in part from the main electricity network; the bakery does not produce any significant waste.

Some questions related to the conceptualisation difficulty are the following. How many business units are there? Two in the legal framework, but one in the statistical framework, which will give different results when producing business demography. Is there a sufficient level of autonomy in the baker's decision-making? If not, how can one obtain the necessary information about the joint decision-making of the baker and his brother-in-law? Which economic activity applies to the village bakery? Granted only one activity per unit, the activities may be baking (bakery) and renting business space (part of the real estate) in the LeU framework, or only baking in the ENT framework. The use of business space activity is not visible in the ENT framework, whereas the retail activity is not visible in either framework, even though it is also value-adding from an economic point of view. Theoretically speaking, should there be 1, 2, 3, or 4 employees listed? Will there be data available to calculate the full-time equivalents?

Moreover, in the broader context, a business generally has a number of relevant properties: location (building), equipment, human workforce, material input, energy input, capital input, output (e.g., product, and service), and waste. Variations that result from some of these being missing in a particular adopted framework can lead to comparability and coherence problems in statistics at the macro-level. Not to mention the other aspects that may be of interest, including factory-less production (no location); owning versus not owning (i.e., renting) the location; outsourcing of work; etc.

8. Dealing with the Unit Problem Under the Total Survey Error Framework

We believe that the unit error should be included and recognised in the Total Survey Error (TSE) framework, alongside the other types of error, such as sampling error, nonresponse error, measurement error, and so on. In other words, while it is important and helpful to try to reduce the unit errors in individual data, it is necessary to approach the unit problem from a more general perspective. It is clear that, in light of the various situations that can lead to unit error, a single-minded focus on the operational aspects of the statistical process will have little effect regarding the conceptualisation and definition errors, and only limited and potentially biasing effects on the observation error.

The effects of the actual unit errors in the collected data, their treatment in data processing and adjustment in statistical estimation need to be understood and articulated under the TSE framework. In terms of data collection and integration, the unit error is rooted in the representation side of TSE framework (Groves et al. 2004; Zhang 2012). However, the different situations of discrepancy that can cause the unit error are interrelated, so that it is important to keep such "interactions" in mind when dealing with the unit problem. Moreover, the unit error will also affect measurement errors and relevance errors on the measurement side of the TSE framework, whereas the causes of

potential errors on the measurement side (e.g., data availability) can also affect one's approach to the unit problem.

9. Evaluation of the Unit Error Through User Value Criteria

Given the lack of established methodology, and the complexity of the unit problem, we believe it will be very helpful to increase our efforts regarding the assessment and documentation of the business unit errors within the international statistical community in order to accumulate and systematically categorise the relevant empirical evidence. Regardless of one's approach to the unit problem, the effects of the unit error that remain in the statistical output can be evaluated with respect to the User Value Criteria, including the so-called quality dimensions, namely, relevance, coherence, comparability, accuracy, timeliness, cost and response burden (Eurostat 2011). All the types of generic unit errors do not affect these aspects to the same extent. For instance, relevance (e.g., output that makes sense to users) seems ultimately rooted in the conceptualisation of the system of statistical units, whereas accuracy (e.g., avoiding bias of various causes) may be more directly related to the observation errors. Hopefully, once the various effects of unit errors have been systematically and amply assessed and documented, one can move on to reform the system of statistical units and the associated regulatory model to develop and implement appropriate methodology in the production and use of business statistics.

10. References

Brion, Ph., T. Deroyon, and E. Gros. 2014. "A First Assessment of Profiling in France." Presentation given at the ENBES workshop "The Unit Problem in Business Statistics", UNECE, Geneva, 10 November 2014. Available at: https://statswiki.unece.org/display/ENBES/ENBES+Workshop%2C+2014+%28Geneva%29+-+The+Unit+Problem+in+Business+Statistics?preview=/126353571/128024595/Unit%20Problem%20Workshop%202014.Brion-Deroyon-Gros.pdf (accessed February 2018).

Delden, A. van. 2017. *Issues when Integrating Data Sets with Different Unit Types*. CBS Discussion Paper 2017-05. Available at: https://www.cbs.nl/en-gb/background/2017/23/issues-when-integrating-data-sets-with-different-unit-types (accessed February 2018).

EEC. 1993. *Council Regulation (EEC) No. 696/93 of 15 March 1993 on the Statistical Units for the Observation and Analysis of the Production System in the Community*. Available at: https://publications.europa.eu/en/publication-detail/-/publication/1ea18a1a-95c2-4922-935c-116d8694cc40/language-en (accessed February 2018).

ESS Task Force on Statistical Units. 2014. "The Statistical Units Model." Written consultation 31 March 2014.

Eurostat. 2011. *European Statistics Code of Practice*. Adopted by the European Statistical System Committee. Available at: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955 (accessed February 2018).

Eurostat. 2015. *Notice of Intention of the Business Statistics Directors Groups (BSDG) and the Directors of Macroeconomic Statistics (DMES) on the Consistent Implementation of Council Regulation (EC) No. 696/93 on statistical units, June 2015*. Available at: https://circabc.europa.eu/w/browse/77baec33-ccf3-409c-a552-9253bd6a9806 (accessed February 2018).

Groves, R.M., F.J. Fowler jr., M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley Interscience.

Haag, O. 2018. "How to Improve the Quality of the Statistics by Combining Different Statistical Units?" In *The Unit Problem and Other Topics in Business Survey Methodology*, edited by B. Lorenc, M. Bavdaz, G. Haraldsen, D. Nedyalkova, P.A. Smith, L-C. Zhang, and T. Zimmermann. Proceedings of the European Establishment Statistics Workshop 2017. Newcastle, UK: Cambridge Scholars Publishing (forthcoming).

Lorenc, B., A. van Delden, P. Struijs, and L.-C. Zhang. 2017. "Statement on the Unit Problem in Business Statistics." Paper written for the European Establishment Statistics Workshop 2017, "Establishment Statistics in a Data-Rich Multinational Environment", 30 August – 1 September, 2017, Southampton, UK. Available at: https://statswiki.unece.org/download/attachments/122325493/ENBES%20Unit%20Problem%20Statement.pdf?version=1&modificationDate=1500370501222&api=v2 (accessed april 2018).

Sturm, R. 2015. "Revised Definitions for Statistical Units – Methodology, Application and User Needs, the Main Conceptual Issues of the "Units Discussion" of the Years 2009–2014." *Statistika* 95(3): 55–63.

Struijs, P. 2016. "The Desired Future System of Statistical Units from the Perspective of Business Statistics." Paper presented at the Fifth International Conference on Establishment Statistics (ICES V), Geneva, Switzerland, 20–23 June, 2016. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/082_ices15Final00136.pdf (accessed February 2018).

Ven, P. van de. 2018. "Economic Statistics: How to Become Lean and Mean?" (JOS special issue on ICES V).

Zhang, L.-C. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica* 66: 41–63.

Arnout van Delden
(corresponding author)
Statistics Netherlands
Department of Process Development and Methodology
Henri Faasdreef 312
P.O. Box 24500, 2490 HA The Hague
Email: a.vandelden@cbs.nl

Boris Lorenc
Bright Lynx Research Consulting OÜ
(At Statistics Sweden at the time of working on the Unit problem statement.)

Peter Struijs
Statistics Netherlands

Li-Chun Zhang
University of Southampton
Statistics Norway