

Statistical Units Delineation and the Quality of Business Statistics

Peter Struijs
Statistics Netherlands
e-mail: psjs@cbs.nl

aim of the paper

1. Business statistics obviously refer to populations of businesses. This implies the application of a definition of the business. Analyses and reports on the quality of business statistics are generally based on sampling theory and often focus on sampling errors, frame errors, non-response errors, classification errors, various types of measurement errors, etc. Frame errors are analysed by looking at causes of possible over- and undercoverage, in which the concept of the business is often taken for granted. However, errors in the application of the definition of the business are not habitually quantified.
2. Most ESS business statistics, such as the structural business statistics, apply the definition of the so-called enterprise, which is defined in the Regulation on Statistical Units¹ as “the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources”. The application of this definition is also called the delineation of the enterprise, since it amounts to determining what legal units out of the universe of legal units belong to any specific enterprise. Unfortunately, the delineation of the enterprise is prone to errors.
3. The enterprise is one of several types of statistical units, albeit the most important one for business statistics. Other types of statistical units, for instance the so-called kind-of-activity unit, also require delineation. Statistical unit delineation has been discussed extensively for many years within and outside the EU, many papers have been written, surveys held, operational rules drafted, and analyses conducted. Nevertheless, despite the existence of a regulation aimed at harmonising the delineation of statistical units, country practices are still quite diverse.
4. As a consequence, the quality of business statistics may be severely affected, in particular in respect of relevance, accuracy, comparability and coherence. However, it is not quite clear in what ways errors in statistical unit delineation may affect the outcomes of business statistics. The aim of this paper is to identify types of errors of statistical unit delineation and their possible effect on the quality of business statistics. The paper gives a logical, qualitative analysis. The reasoning does not depend on the specifics of any actual statistical unit definition.

statistical unit delineation

5. Why is statistical unit delineation needed at all? In fact, National Statistical Institutes (NSIs) that spend significant resources in the delineation of units have not always done so, and there are many NSIs that even now do not take the effort. If units are not delineated by the NSI, the units that are registered by

¹ 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.

other institutions are the “businesses” on which the NSI builds its statistics. Examples of such unit types are the so-called legal units from chambers of commerce or other official registrations, and tax-paying units registered by the fiscal authorities. In fact, there are many types of administrative units. Using administrative units for business statistics is low-cost and may be easy.

6. The use of administrative units has important drawbacks. Countries that started delineating statistical units before this was regulated at the EU level observed that some companies created daughter companies for all kinds of reasons, such as minimising taxation, eligibility for subsidies, spreading or isolating risks, reducing liability, preparation of family succession, a better administration of delegated managerial responsibilities, or even obscuring the effective ownership of patents to competitors. Companies combined into groups, holding companies emerged, etc. As a consequence, a real business, economically speaking, could consist of a number of administrative units. For such cases, statistics about populations of legal units or other types of administrative units would not show the real economic picture. The relevance of business statistics was at stake.
7. Moreover, if a business consists of several legal units that are together run as one single economic entity, its own accounts would reflect that. This makes data collection from and about separate administrative units more difficult and less meaningful, especially if the data registered at the level of the legal unit is artificially deconsolidated data, not registered for business purposes.
8. Such considerations of relevance and data obtainability triggered some NSIs to define their own units for business statistics, such as the statistical enterprise, as distinct from administrative types of units, such as the legal unit. Thus, statistical unit delineation became necessary. In many instances administrative units could still be used without any modification for statistical purposes, but in the case of constructs of administrative units, delineation would be needed. However, unit delineation requires significant resources and methods for unit delineation were not readily available, which kept many NSIs from introducing such practices.
9. At the EU level, the need for defining statistical units was also recognised. An additional consideration at the EU level was that the administrative units in its member states were defined at the national level and lacked comparability. And to the extent that NSIs actually delineated statistical units, practices were quite diverse. At the same time, as the system of European statistics developed, the need for applying different statistical unit types for different types of statistics became clear. This led in 1993 to the adoption of the Regulation on Statistical Units, which is still in place.

errors in statistical unit delineation

10. Suppose an NSI tries to apply the definition of a statistical unit type, such as the enterprise, and let us ignore the international dimension, that is, let us assume that the unit to be delineated is known to consist of only national administrative units, or parts thereof. Why should errors occur at all? An important reason is that statistical units are defined for statistical purposes and administrative units for administrative purposes, so that errors can only be avoided if all administrative units are checked for their suitability for statistical purposes. Unless this can be done in an automated way, it is not feasible to check all units.
11. Automated as well as manual delineation of units requires two things: a set of operational delineation rules and information about the real world situation of the units for these rules to act on. This information is often not available without collection effort. This means that if delineation is automated, errors inevitably occur. Automated delineation may result in decent proxies of statistical units and may be

efficient, so may be used anyway. NSIs often limit manual delineation to the largest and most complex units, where they cannot afford errors. This is called profiling.

12. Definitions of statistical units need interpretation. For instance, for the enterprise, operational rules are needed for interpreting what it means to “benefit from a certain degree of autonomy in decision-making”. In fact, such definitions may be the result of negotiations between NSIs and Eurostat, all having their own interests and objectives, leading to rather vague formulations that are acceptable to most. Operational rules may also be the result of negotiations and may still leave room for different interpretations. This year, after 22 years, the ESS directors of business statistics and macroeconomic statistics finally agreed on operational rules for the units of the Regulation on Statistical Units². However, these rules do not cover all situations encountered in practice. Nevertheless, the operational rules provide good guidance.
13. In this situation, what qualifies as a unit delineation error? To the extent that the definitions and operational rules are clear, not applying them would imply a unit delineation error. This means that at least one administrative unit, or part thereof, is incorrectly not included in the set of administrative units that together form the statistical unit, or is incorrectly excluded from that set. That is, the coverage of the statistical unit in terms of administrative units is incorrect. However, in terms of variables such as employment or capital, the error may be fussier, as the real-world constructs of combined administrative units also pertain to those variables.
14. Unit delineation errors may be caused by giving a wrong time-stamp to changes. For instance, if two legal units, each equaling an enterprise, merge into one enterprise, this may be known to the NSI only with some delay. In the meantime, the two enterprises are wrongly kept apart by the NSI, constituting a delineation error (or two errors, depending on the point of reference). As mergers are often the result of a long process, at some moment passing the point of no return, the timing of the change may contain an element of arbitrariness.
15. There are some parallels between unit delineation errors and errors of classification to industrial activity. Such classification systems also contain definitions and operational rules, and there may always be situations for which those rules do not readily provide an answer. In addition, classifying all units manually is not feasible, and classification errors cannot be entirely avoided. And when a statistical unit is classified into another class, the timing may also contain an element of arbitrariness.

effects of errors in statistical unit delineation

16. What are the effects of unit delineation errors on business statistics? Let us first consider the situation of a closed universe: the set of administrative units that form the basis of the business statistics considered is complete, only their allocation to the statistical units of the business population may contain errors, so there is no interchange with populations outside the business statistics considered. Unit delineation errors then imply that statistical units have an incorrect coverage of administrative units, but the incorrect allocation of administrative units remains within the business statistics concerned.

² Notice of intention of the Business Statistics Directors Groups and the Directors of Macroeconomic Statistics on the consistent implementation of Council Regulation (EC) No 696/93 on statistical units, June 2015.

17. At the level of such an erroneously delineated statistical unit itself, there will likely be incorrect unit characteristics such as size in terms of employment as well. The industrial activity code may be wrong, but this is not necessarily the case. If an administrative unit is incorrectly not comprised by a statistical unit, it will be classified according to its own activity, which may very well be different from the activity of the statistical unit with which it should have been combined, but the code of that statistical unit may not be affected. Apart from errors in unit characteristics, other variables may obviously be influenced as well. Derived variables such as turnover per employee may also be affected.
18. At the level of the business statistics as a whole, the closed universe, there may be several effects, that is, errors in the published aggregate results. First of all, the number of statistical units will be wrong. If administrative units are wrongly considered statistical units in cases where they should have been grouped into statistical units, the number of units in the statistics will be too high. As their sizes contain similarly biased errors, the average unit size will in reality be higher than published. There will also likely be errors in the distribution of the total population in terms of unit characteristics, such as industrial activity, but some errors may cancel each other out.
19. As to the reported statistics on variables other than unit characteristics, two types of such variables may be distinguished: those that are additive when administrative units are consolidated, and those that are not, i.e. that are sensitive to consolidation. For instance, the volume of employment (if counted in terms of hours, not number of persons employed) is not sensitive to consolidation, that is, the employment concerned is allocated to the wrong statistical unit but the total employment remains the same. Contrarily, turnover is consolidation sensitive, that is, if the output of an administrative unit is the input of another administrative unit, their combined turnover will shrink if the two are combined into one statistical unit.
20. If we look at a subpopulation of business statistics, for which the closed universe assumption does not hold, the situation is different. The subpopulation may have under- or overcoverage in terms of administrative units. As unit delineation often involves including ancillary activities or vertically integrated services in statistical units, the size of services industries tend to be overestimated, and the size of manufacturing industries underestimated, if an NSI does not apply unit delineation. Business statistics with a lower cut-off, that is, which cover only units with a minimum size, may be particularly affected by unit delineation errors, as such errors tend to change the size distribution of the population.
21. Business statistics are usually produced as part of time series. Unit delineation errors will also influence the reported evolution of industries. This effect may be small, if the structure of administrative and statistical units is stable, but even then the allocation of aggregate change to industries may be consistently erroneous, that is, biased. If structural changes do regularly occur, the effect of delineation errors may be worse and may also comprise errors in the reported period of occurrence of changes.
22. The unit delineation errors described so far have a clear effect on the quality aspects of relevance, accuracy and coherence. However, the main quality aspect that is negatively

influenced is probably international comparability. If one NSI applies unit delineation and another uses administrative units as statistical units, the comparability of their business size distribution and of the relative size of industries may particularly suffer.

23. Finally, there may be important indirect effects of unit delineation errors. The subpopulation to which a delineated unit belongs, in particular its industrial activity and size class, may determine its possible inclusion in a survey. If sampling is done on the population of administrative units rather than on the population of statistical units, unit delineation errors will only have an effect at the stage of post stratification, but if samples are drawn from statistical units, unit delineation errors will have to be taken into account in the sampling design, in theory necessitating oversampling.

the way forward

24. There are several possibilities to reduce the negative effects of unit delineation errors to users of business statistics. First of all, as long as delineation practices remain as they are, transparency in quality reports would serve users well. Second, since there are now agreed operational rules for statistical unit delineation, these should be applied by all NSIs of the ESS. In addition to these operational rules, the directors concerned adopted a set of guiding principles for their application, in which important aspects such as subsidiarity and proportionality are discussed. In particular profiling of the largest businesses is recommended.
25. Already before the adoption of the Regulation on Statistical Units, businesses started to operate increasingly internationally, and there are many large multinational enterprises. This phenomenon is not adequately dealt with in the regulation, and neither in the operational rules. However, without having reached the stage of formal legislation, the ESS has worked on this issue, among other things by designing a European Group Register (EGR). If all NSIs of the ESS delineate their statistical units in a way consistent with the EGR, this would enhance international comparability.
26. Finally, systematic quality management of business statistics should incorporate quality aspects of unit delineation. Setting such quality norms, derived from minimal output quality requirements, and measuring the degree of adherence would be an important step forward. As unit delineation errors are not independent from industrial classification errors, and the nature of these domains show some parallels, it is recommended to integrate the quality management of these domains. It is also recommended to continue with the further development of operational rules, in order to reduce the room for different interpretations of statistical unit definitions.