

Book Review

*Peter Struijs*¹

Boris Lorenc, Paul A. Smith, Mojca Bavdaž, Gustav Haraldsen, Desislava Nedyalkova, Li-Chun Zhang and Thomas Zimmermann, eds. *The Unit Problem and Other Current Topics in Business Survey Methodology*. 2018 Newcastle upon Tyne: Cambridge Scholars Publishing, ISBN 978-1-5275-1661-8, 288 pages.

In business survey methodology, increasing attention is given to the “unit problem”. The unit problem is the set of issues associated with the application of the statistical concept of the unit in business statistics. Various types of units may be defined and applied by statisticians, administrative registers and businesses themselves, which greatly complicates the production of statistics. Issues include the delineation of units, for instance if an enterprise is composed of several legal units. They include relating collection and observation units to the target units of business statistics, for instance if local units are used for collecting data on businesses. They also include identifying the units to be listed in business registers, especially if different types of units are used in business statistics, as is the case in the European Statistical System (ESS). Next to the enterprise, which is defined as an independent business actor in the economy, the ESS makes use of, for instance, the kind-of-activity unit, which is roughly defined as part of an enterprise carrying out a single industrial activity and administered as an entity in its own right. There are associated issues of sampling methodology, of coordination and integration of business statistics, of costs, efficiency and data availability, and more. Furthermore, all issues can be seen from the perspective of the errors they may cause in the statistics produced.

Given the high relevance of the unit problem, the book *The Unit Problem and Other Current Topics in Business Survey Methodology* by Boris Lorenc *et al.* (eds) is a very welcome contribution to understanding such issues. However, the book, which comprises an edited and enriched selection of papers presented at the 2017 European Establishment Statistics Workshop, also covers other business survey topics, related or less related. Examples include sampling coordination, managing response burden, questionnaire design, using new data sources for price statistics, and data visualization. The book will not only benefit survey methodologists and producers of business statistics in general, but also the users of such statistics, widening their understanding of business statistics and their intricacies.

Roughly half the book is concerned with the unit problem. The unit problem itself is explained in a chapter by Smith, Lorenc and van Delden, in which they consider unit errors in the context of the five dimensions of output quality as applied in the ESS.

¹ Statistics Netherlands, CBS weg, Heerlen, 6401 CZ. Netherlands. Email: p.struijs@cbs.nl

These are user relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, and coherence and comparability. Such a quality framework is helpful in dealing with the unit problem in a structured way. Sturm discusses the unit problem in the context of business registers in Germany, where profiling (i.e., unit delineation) is being implemented. Important issues are the relationship between the enterprise and the kind-of-activity unit as defined and applied in the ESS, and the definition of the enterprise itself. The discussion in the ESS on these matters has not yet been settled.

The book includes several French contributions touching on various aspects of the unit problem. Haag explains how the French business register (actually a network of registers) makes a distinction between enterprises and legal units. The latter used to be the unit applied in business statistics, but since 2010 a change to using the enterprise has been carried through, as that unit has more economic relevance, especially if business structures are complex. Haag quantifies the differences between enterprise and legal unit populations, thereby illustrating the importance of making the distinction. Two-stage cluster sampling is applied to deal with the difference between the collection unit and the statistical unit, as explained by Gros and Le Gleut. The design optimizes the statistics produced at the level of enterprises under a constraint on the number of legal units surveyed. A second contribution of Gros and Le Gleut explains how samples for different surveys are coordinated, positively and negatively, using permanent random numbers. The transition from legal units to enterprises as the unit of business statistics has consequences for the treatment of influential values, that is, winsorization. This is the topic of a contribution by Fizzala.

Four more chapters are clearly linked to the unit problem. Van Delden looks at issues arising when statistics are based on integration of various data sources. These may have different unit types. By using a linkage and data integration framework, he identifies the issues that occur in twelve Dutch case studies. Lammers discusses ways to improve the efficiency of profiling in the Dutch business register. The analyses of the process are based on process mining techniques, which involve the generation of mostly quantitative metadata from the applications used. The unit problem also has a regional dimension, which Ichim discusses for the Italian case of producing business indicators using territorial domains. This requires a strategy for linking enterprises to the local unit level. An analysis of the response process for Norwegian business statistics is provided by Haraldsen, in which the risks to data quality are central. These risks are related to the complexity of the unit structure, which also largely determines the response burden.

Concerning the other topics of the book, two chapters are about sampling. Zimmermann, Schmiedel and Lorentz show how the German federal statistical institute responded to a court decision requiring a much more even spread of the response burden for the services sector. A new sampling design was developed, but take-all strata could not entirely be avoided. The Dutch methodology of sampling coordination across all business surveys and panels is described in a contribution by Smeets and Boonstra. Use is made of permanent random numbers and response burden values. Two more chapters concern data collection. For the United Kingdom, Steward, Sidney and Timm looked at paradata generated during the completion of questionnaires in order to improve them. Both quantitative paradata was used, such as the number of error messages, and qualitative paradata, such as call records. Distinguishing between different respondent types is key to

their analysis. A Swedish experiment with validation embedded in an electronic data collection tool is the subject of a chapter by Lorenc, Norberg and Ohlsson. The experiment sheds not only light on data quality aspects, but also on the relationship between data validation, response burden and costs.

Another topic of the book is the change in the use of data sources for price statistics. This topic is introduced by Zhang, who mentions the increased use of web scraping and scanner data, in addition to the more traditional data collection through enterprise reporting. As a consequence of the change, issues arise concerning the index formula to be used. A Canadian investigation of the possible use of scanner data and the challenges posed is described by Deshaies-Moreault, Harper and Yung. The current CPI approach is compared to one in which scanner data plays a more prominent role. The Slovenian approach to price surveys is explained in a contribution by Razinger, including new developments concerning data collection methods. The last topic of the book, which is also relevant outside the realm of business statistics, is data visualization. Vila, Cervera-Ferri, Camões, Bolko and Bavdaž make a case for its relevance to statistics, and argue that good visualization requires a worked-out methodology in which the cognitive processes underlying data interpretation play a central role. This can be studied empirically, for instance by studying alternative presentations and using eye tracking and gamification.

The Unit Problem and Other Current Topics in Business Survey Methodology is a rich book that is accessible and worth reading by all interested in business survey methodology, including the users of business statistics. Moreover, its special focus on the unit problem and its consequences for business statistics makes it unique.