

What makes business statistics different?

Wim Kloek, Eurostat¹ (wilhelmus.kloek@ec.europa.eu)

Paper prepared for the European Establishment Statistics Workshop (draft 11.09.2011)

Neuchâtel, 12 – 14 September 2011

Introduction

A natural way to cluster statistics is on the basis of the statistical units, the carriers of the characteristics we measure. The two most prominent clusters are business statistics and person and household statistics; other statistical units are location and product. Integration of data inside the cluster is natural. The common sampling frame offers the basis for integration. The clusters are also connected. For instance employment links business statistics to person and household statistics. Employment is a resource and a cost to businesses; to persons it is a category of time use and a source of income. There are also connections via other types of statistical units. Products are the result of economic activities by businesses and can be consumed by persons. In the scope of this article the exact number of clusters and their relations are not important, what concerns us is the question in what sense the clusters require different statistical methods. In what follows I will concentrate on the comparison between business statistics and person and household statistics.

In what follows I will briefly discuss three fundamental reasons for different methods in business statistics and person and household statistics. Along the statistical business process steps I will try to identify the potential impact on statistical methods. On the basis of this I will reflect on potential consequences on the organisation of methodological development and support and organisation of statistical institutes.

Language

Everyone who has migrated between business statistics and person and household statistics will have experienced a cultural difference also reflected in language. For instance, the statistical business register is a kind of sampling frame, a form a kind of questionnaire and an observational unit a respondent. This pattern is true over several languages. Language in business statistics is closer to accounting, whereas in person and household statistics the language of survey methodology is used. Awareness of this is essential for communication between the clusters. For the purpose of this paper we should ignore differences in language and look for the real differences.

¹ The contents of this paper do not necessarily reflect the opinion or position of the European Commission.

Three differences

In my understanding there are three core differences that might justify differences in methods:

- Statistical units
- Skewed distributions
- Character of variables

Statisticians might discuss which persons should be included in the survey population, but they seldom discuss the definition and borderlines of a person. This is already different for households. Conventional criteria are needed to clarify to which household students, prisoners, persons hospitalised etc. belong. Businesses can have complex structures: x legal entities with y economic activities on z locations. Although some guiding principles exist on the delineation, the result will always depend on the national legal and fiscal system, on administrative procedures and on subjective judgement. The complexity of statistical units has an impact on

- harmonisation of statistics over countries (application of common profiling rules, coherent treatment of multinational enterprises);
- sharing the understanding on the delineation with the respondents and data users;
- spread of information within businesses (different administrative systems for different purposes).

The high level of complexity is not the average case; many businesses consist in a sole proprietor with activities on just one location. However, the complex business are usually more significant in economic terms. Put differently, the problem of complexity of statistical units is correlated with the skewness of distributions. Most of the core target variables in business statistics (value added, turnover, exports etc) have a skewed distribution. Core variables in person and household statistics are often of qualitative nature (employment, housing, health etc) and some quantitative variables have an inherent limit (e.g. time use) which prevents very skewed distributions. Some skewed distributions do exist, however, for instance the distributions of income and wealth.

The character of the variables is the third difference; it goes beyond differences in format (quantitative/qualitative). Most key variables in business statistics have their basis in the accounting system. This has two important consequences:

- terms and definitions are embedded in the accounting tradition; this will make it difficult to communicate statistical terms and definitions that differ from this tradition;
- the data should not be regarded as separate pieces of information, but as integrated in balance sheets.

Information on the stratifying classifications (economic activity, size class, region etc) is usually not collected in the business survey process, but are taken from the statistical business register. This procedure supports the coherence of business statistics. Especially the coding by economic activity (NACE) is a significant example of a non-accounting variable; it is applied in all business statistics to delineate the population and breakdown in subpopulations. Information collected from person and household statistics is less pre-defined and less structured. It has the advantage of different loosely connected concepts, thus spreading the risk of misinterpretation and irrelevance and creating a much more open and flexible system. The downside is that the system of variables has no natural starting point and that the concepts remain under continuous discussion. Most variables try to capture facts, like the number of visits to the dentist, the type of house or the number of hours worked. However, to capture aspects of well-being also the measurement of subjective variables is relevant: feelings, perceptions, opportunities.

We have seen that the three differences between business statistics and person and household statistics are of a gradual nature, the importance depends on the concrete case under consideration. We have also seen that the three differences tend to be correlated.

Different methods required?

The subsequent question is whether the differences indicated before require different statistical methods. I will discuss the methods along the statistical production process. For this purpose I have stolen an overview of statistical methods along the statistical process from the currently ongoing ESSnet project Methodology for Modern Business Statistics ([Memobust](#)). The intention of this project is to draw up a comprehensive overview of methods relevant to European business statistics; the overview was taken from an internal document. The overview is linked to the Generic Statistical Business Process Model ([GSBPM](#)), but allows a somewhat different view probably more fitted for classification of methods.

I offer my apologies for further simplifications and improvements.

The overview is used here to identify where a similar handbook on person and household statistics would be different. I distinguish three cases:

NO: No significant changes required

YES: Significant changes required

NEW: The relevant items will be completely different

My preliminary classification could sometimes disclose lack of knowledge on certain methods; the reader is invited to contribute to the classification. The comment is intended to justify the choice. It will require the involvement of many others to get a more complete overview and to validate the result.

Are different methods required in business statistics?

Topic	Comment	Change?	GSBPM
User Needs	Respondent as user	NO	1
Overall Design	Large businesses: longitudinal data	NO	2
Improving repeated surveys	Use of process information	NO	2
Design of Statistical Concepts	Requires solid knowledge of accounting	YES	2
Questionnaire Design	Statistical units, internal organisation	YES	3
Pilot Surveys	At location: internal organisation	NO	3
Frames	The delineation and classification of units in the business register is a considerably more difficult task; skewed distribution and potential impact of frame errors	NEW	4
Sample Selection	Skewed distribution	YES	4
Collection of Secondary Data	I guess only few generic methods anyway	YES	4
Data Collection	Special strategy for large businesses	YES	4
Mixed Mode	Probably different mode effects: soft variables could be more sensitive	YES	4
Response	Cost model more relevant for business statistics	NO	4
Micro Integration		NO	5
Coding		NO	5
Error detection	Skewed distributions: selective detection; soft variables: less rules	YES	5
Error correction	Inaccurate for small number of large units	YES	5
Derivation		NO	5
Weighting	Higher risk with frame errors (outliers)	YES	5
Estimation	Non-response in large units; small area estimation skewed distribution and different units	YES	6
Macro Integration	Special attention for large businesses	YES	6
Time-series Analysis	Reclassification of large businesses can produce discontinuities; difficult to interpret for soft variables	NO	6
Disclosure Control	Tabular data: large businesses Person and household: micro data	YES	6
Dissemination		NO	7
Documentation		NO	8
Evaluation		NO	9
Quality		NO	overall

From this overview I conclude:

- that the role of the statistical business register in business statistics is more significant and more complex than the role of sampling frames in person and household statistics;
- that special attention should be paid to large businesses throughout the production process;
- that the soft variables in person and household statistics make it more difficult to validate, analyse and interpret.

In order to derive more significant conclusions it would be good to add more dimensions to the classification of differences. For instance: method, implementation, infrastructure.

Reflection on organisation

What could be the consequences of these conclusions on the organisation of statistical institutes and on the organisation of methodological services within statistical institutes? To begin with the last question, I would argue that once the differences between business statistics and persons and household statistics are well understood, methodological support could better be centralised for the following reasons:

- Much of the methods are common to both clusters;
- The differences are gradual, not absolute; statistics on business start-ups are more similar to person and household statistics, whereas the household budget survey has more in common with business statistics;
- Centralised methodological support is a natural ally in standardisation and integration, thus supporting quality and efficiency.

As the crucial role of the statistical business register is usually not well covered in academic education, junior methodologists should profit from an on the job training in the business register as soon as possible.

Statistical institutes can be organised by statistical domain or by business process step. Both approaches are valid. The advantages of organisation by statistical domains are:

- Total process management improves the quality of validation and interpretation of results, because only knowledge of all process steps can help distinguish between real developments and artefacts produced in the production process;
- It is rewarding, as it connects everyone directly to the statistical output.

Please note that the use of administrative sources seriously hampers total process management.

The advantages of organisation by business process step are:

- Support of standardisation of methods, tools and processes;
- Support of integration of statistical information, thus improving coherence of statistical information and producing new output.

Organisation by statistical domain is dominant in the European statistical institutes. This would urgently call for a shift to organisation according to business process step in order to develop the advantages of standardisation and integration.

As both organisational principles are valid and have their own benefits, reorganising once in seven years according to the other principle would avoid one-sided development. The optimal solution would be to somehow reconcile both approaches.