

Survey frame methodology

Ph. Brion

ESCWA workshop - Amman

26/29 September 2016

Outlines of the presentation

- Introduction
- 1. Identify concepts
- 2. Design outputs
- 3. Design frame and sample
- 4. Design collection

Introduction (1)

- This presentation is based on chapter 8 of the Guidelines on Statistical Business Registers
- This chapter takes the GSBPM (Generic Statistical Business Process Model) as a basis to focus on some sub-processes :
 - Identify concepts
 - Design outputs
 - Design frame and sample
 - Design collection

Introduction (2)

- This presentation will also use some examples from the French system (even if each context is different)
- In France : an inter-administrative business register, SIRENE (legal units and local legal units), with an id-number used by all administrations
- ... and, since 2012, a statistical business register, SIRUS, that is now the backbone for economic surveys

1. Identify concepts

1. Identify concepts (1)

- This part of the chapter is mainly focussed on the issues of frames that have to be, as much as possible, close to the population under review
- For example, concerning short-term statistics frames have to take into account the entries and exits of the register
- And, for statistics referring to the same period, a single frozen frame should be used to give coherence to the statistical production

1. Identify concepts (2)

- If these principles seem obvious, their practical application is not simple, because of the impossibility to have a continuous updating of the register
- The idea is, then, to get, as much as possible, continuous information for the updating, and also to be able to know, for a given period, what is the reference population
- Example : SIRUS, the French statistical business register, is updated every day by SIRENE, concerning the creations of legal and local units
- ... but not all variables are updated in a continuous way

1. Identify concepts (3)

- Example (continued) : in SIRUS, different kinds of updating :
 - Mass updates
 - Feedback from some statistical surveys
 - Specific updating operations
- Also, beside the updated value, three kinds of metadata are available for the main characteristics :
 - Reference date of the event (for example for a death)
 - Date of the updating
 - Source of the updating

2. Design outputs

2. Design outputs (1)

- The statistical business register is used to provide survey frames, which means :
 - A population of interest
 - Identification variables (including the address)
 - and variables to be used for the sampling plans (to be discussed with business statisticians)

2. Design outputs (2)

- The question of the populations of interest raises the more general question of the scope of the business register (see also next part)
- Example (SIRUS): SIRUS does not include all units belonging to SIRENE, the inter-administrative business register, since some units are registered only for administrative needs
 - For example, some units are registered in SIRENE, for real estate management by the different members of a family in case of inheritance of a building, and not included in SIRUS

3. Design frame and sample

3. Design frame and sample

- Choice of statistical unit
- Target populations, frames and coverage
- Design frame methodology
- Over-coverage and under-coverage errors
- Misclassification by principal economic activity
- Dealing with frame errors
- Sample design

3.1. Statistical units (1)

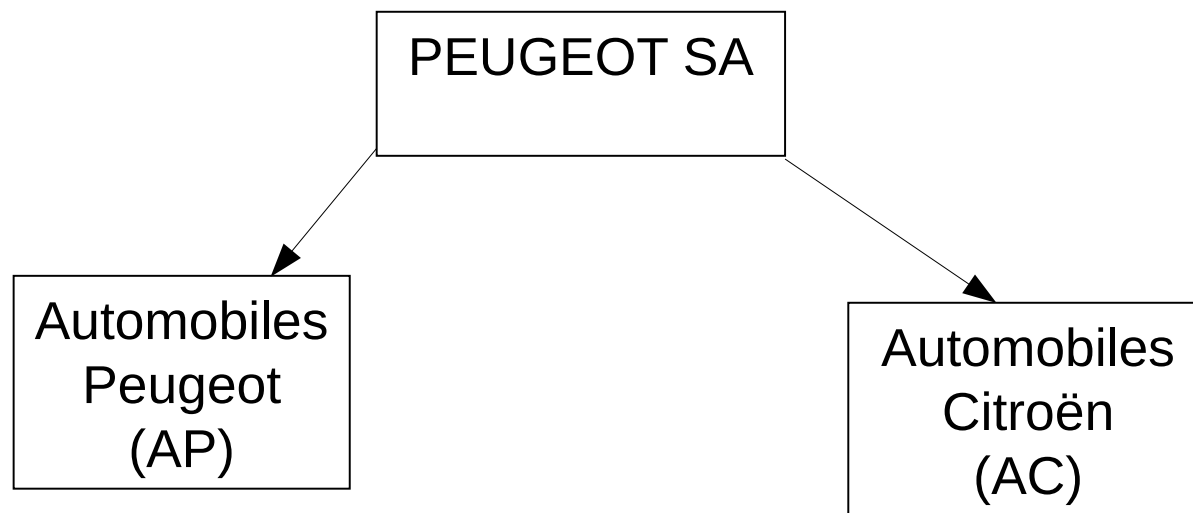
- Different statistical units in the business register
- The use of a unit depends on the topic of a statistical process
- Types of statistical units :
 - Legal unit
 - Local unit
 - Enterprise
 - Enterprise group
 - KAU
 - Establishment

3.1. Statistical units (2)

- Example (SIRUS) : SIRUS was build using first SIRENE (with legal and local units), but two « new » units were introduced :
 - The group
 - The enterprise (referring to the European definition)
- For the group, use of the LIFI system
- Two ways of building the enterprises :
 - For the largest groups, manual profiling (with face-to-face contact with him)
 - For the other ones, automatic profiling from legal units

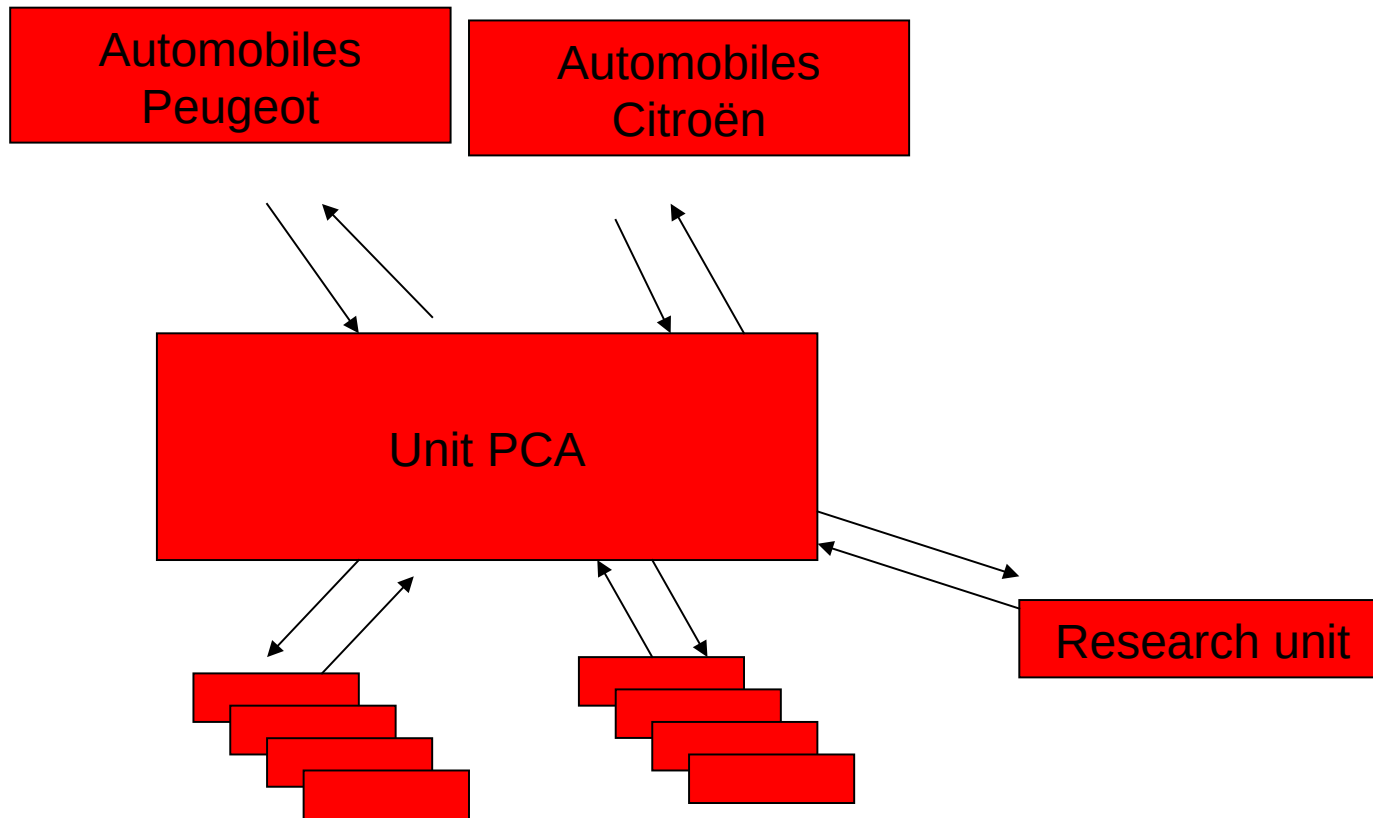
3.1. Statistical units : the example of Peugeot (1)

- Before the reorganization (end of the 90's) : two legal units did produce and sell cars
- The factories were not legal units



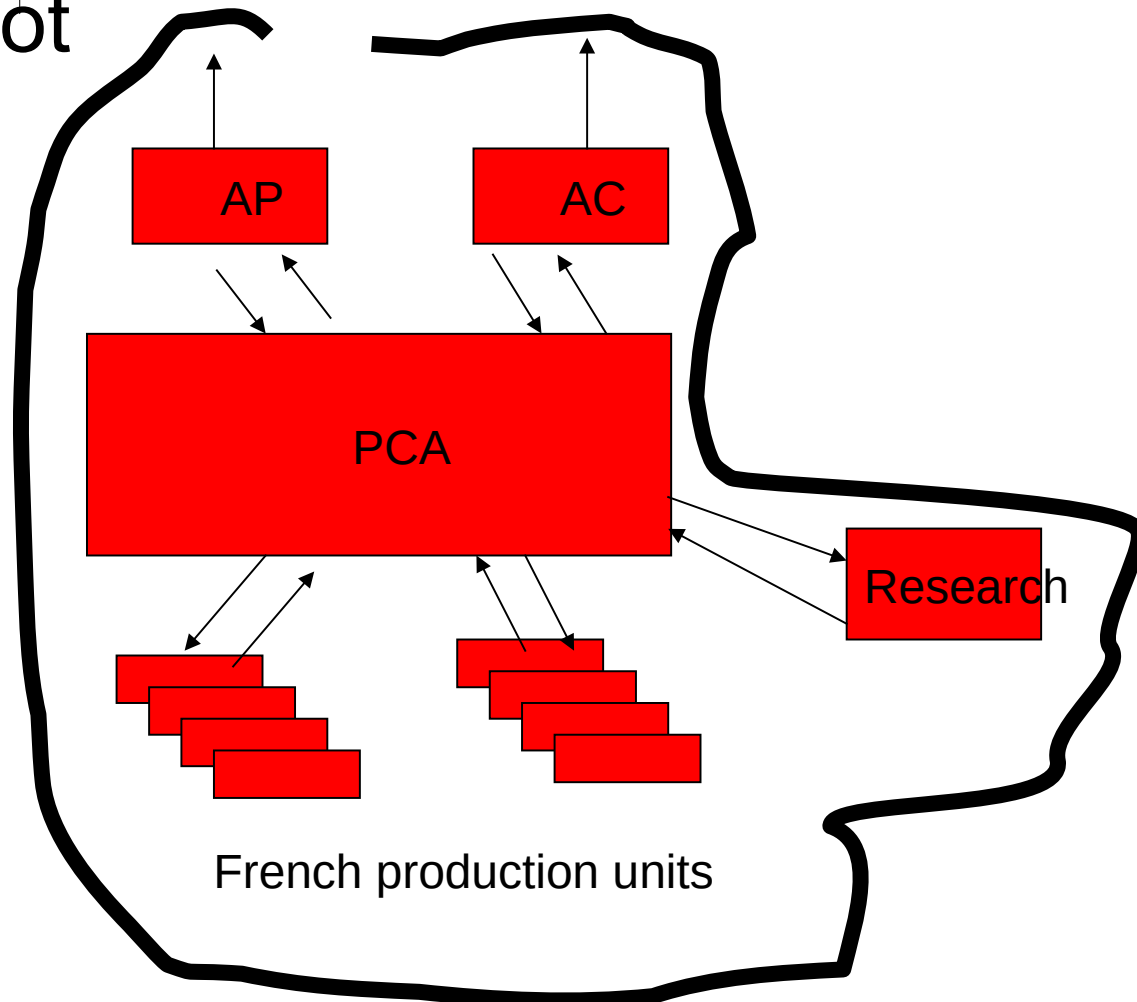
3.1. Statistical units : the example of Peugeot (2)

- After the reorganization

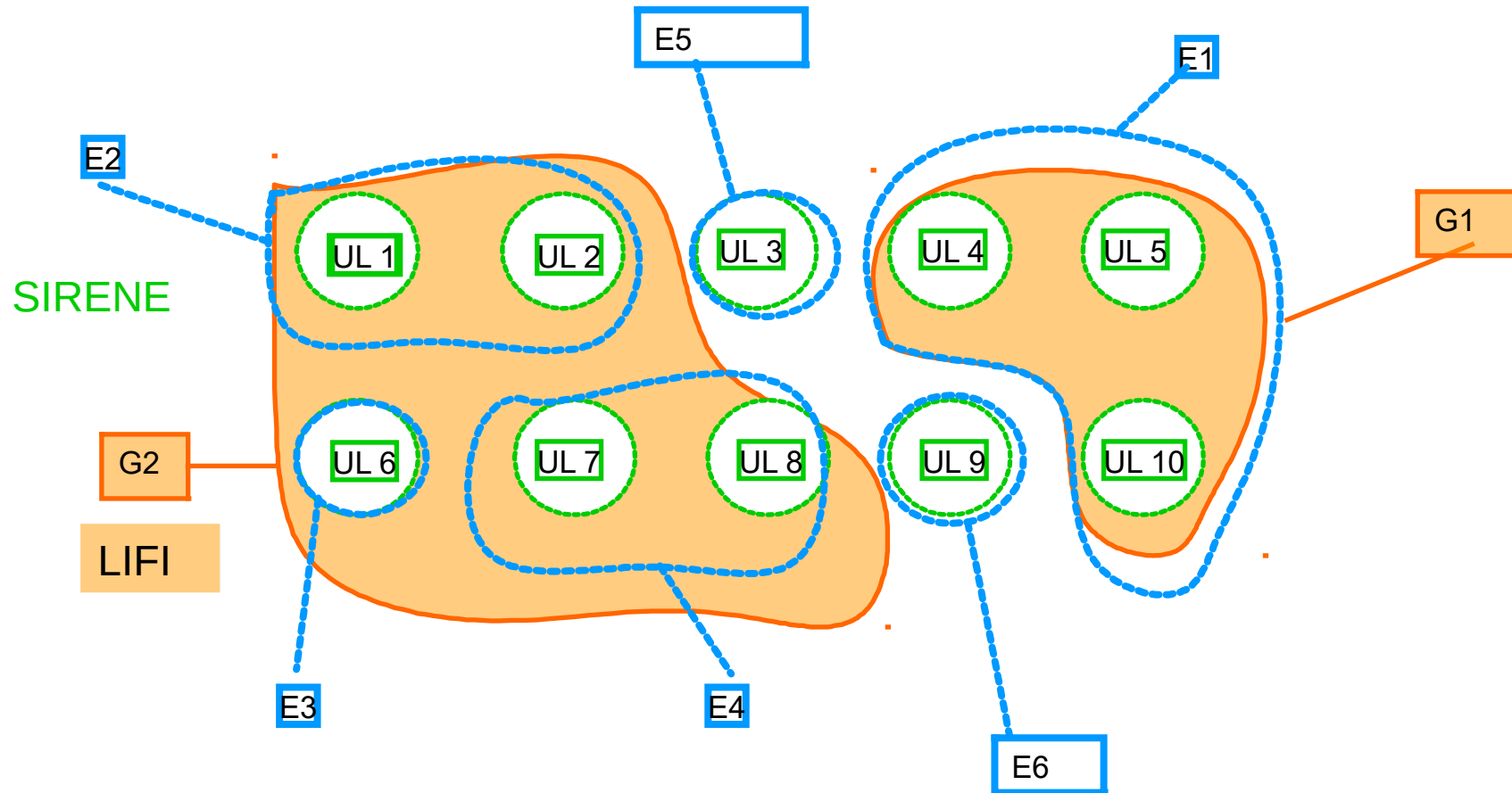


3.1. Statistical units : the example of Peugeot (3)

- Creation of a unit « enterprise » within the group Peugeot

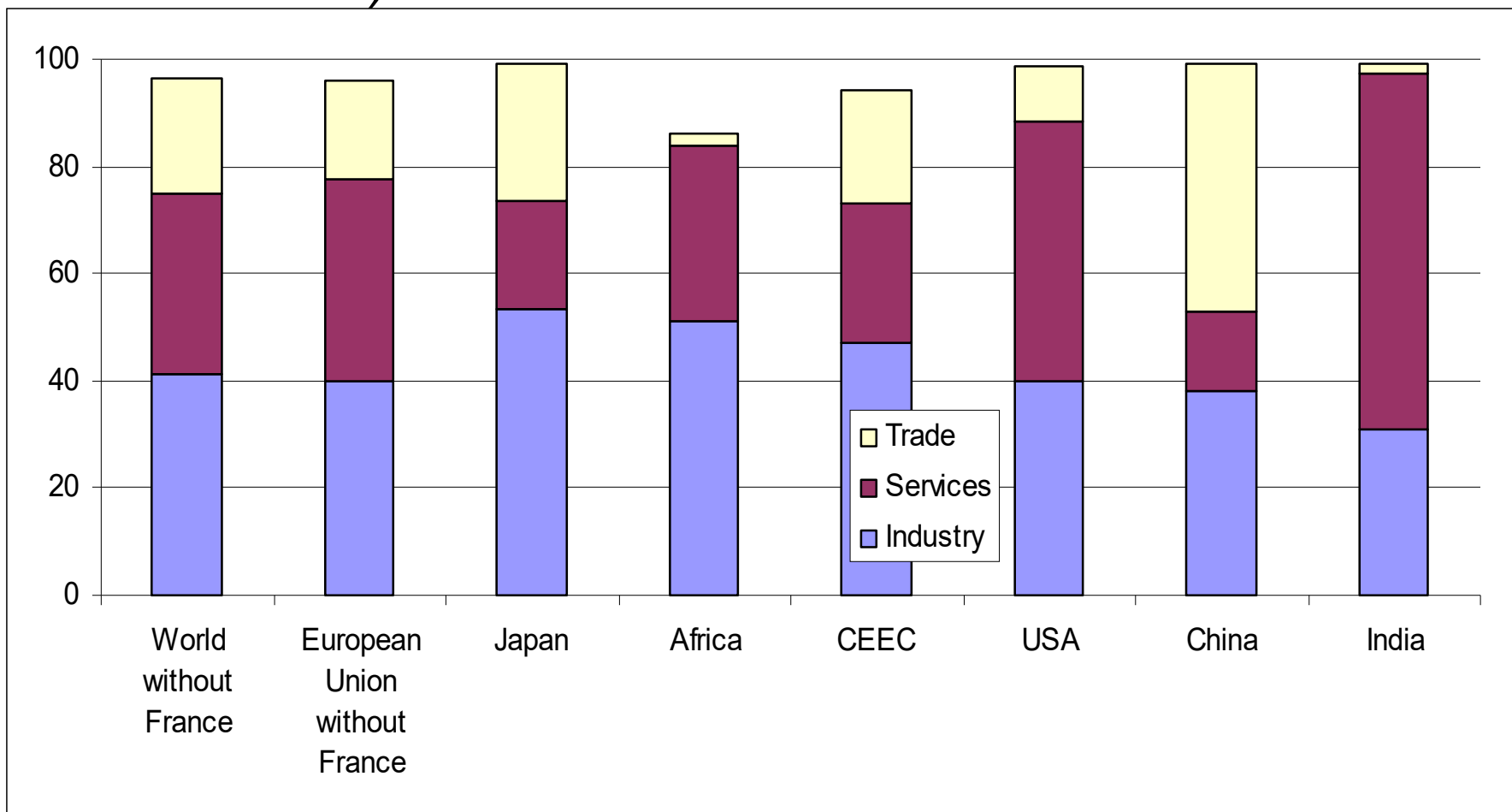


3.1. The statistical units in SIRUS : legal units, groups, enterprises



3.1. An example : the use of the group for the OFATS survey

- OFATS ; survey of the affiliates « abroad » of the French groups (*here, breakdown of the employees - OFATS 2010*)



3.2. Target populations, frames and coverage (1)

- The frame population is the best approximation of the target population, using characteristics available in the register (economic activity, size class, ...) to build it

3.2. Target populations, frames and coverage (2)

- Example (SIRUS): the « main activity » and the « juridical category » are the basic material to define the scope of the productive sector (for example for structural business statistics)
 - However, some adjustments are necessary (example of societies created just for the management of buildings, and mixed, concerning the juridical category, with the family societies created in case of inheritance of a house) : in some cases, units are included in the scope referring to the fact they have a fiscal declaration

3.3. Design frame methodology (1)

- Extraction, at a specified moment, of a snapshot of the register, with units being active
- The survey frame generally consists of one set of statistical units (for example enterprises, or establishments), and occasionally may be designed using two types of units

3.3. Design frame methodology (2)

- The time for the extraction depends on the strategy of the survey (calendar)
- For short-term statistics, if possible introduction of newly active units each month / quarter, and deactivation of ceased units
- Example (SIRUS) : the metadata about the reference date of an update are very useful for building sampling frames corresponding to the target populations (the reference periods being different, for example, for short-term and annual statistics)

3.4. Over-coverage and under-coverage errors (1)

- Over-coverage mainly for two reasons :
 - Inactive units classified as active in the register
 - Duplicates of units
- Under-coverage :
 - Sources used to create and maintain the register are not complete (due to time lag, to thresholds, to units not included, etc.)

3.4. Over-coverage and under-coverage errors (2)

- Reduce over and under-coverage by using information from other sources, if possible
- If over-coverage can be detected during the survey process and does not introduce bias (but if the over-coverage is important, it increases the variance of the estimates), under-coverage introduces biases in statistical estimates :
 - Some form of compensation must be introduced, by statistical weighting, or extrapolation from other sources

3.5. Misclassification by principal economic activity (1)

- The misclassification of some units is unavoidable, since it is impossible to update the code of each unit
- It can introduce biases in case of use of a subset of the frame corresponding to specific industries
- Some checking operations may be undertaken, if one knows which parts of the register are the most concerned

3.5. Misclassification by principal economic activity (2)

- The feedback of the surveys may be used to make a (partial) updating of the register, but also through the estimates
- Example (INSEE) : for the results of the French structural business statistics, the misclassifications that have been observed with the statistical survey are used through statistical estimates to make inference for the whole population (*reference : Brion, Gros (2015) Statistical Estimators Using Jointly Administrative and Survey Data to Produce French Structural Business Statistics, Journal of Official Statistics*)

3.6. Dealing with frame errors

- The errors that have been detected have to be corrected, but their correction may have consequences, particularly on the level of some aggregates (and, as a consequence, on evolutions)
- It may be interesting to produce two kinds of results :
 - With / without the correction of the frame
- Generally, since large enterprises are included in survey samples, and then updated, the impact of errors is not significant

3.7. Sample design (1)

- The efficiency of the sampling plan will be closely linked to the « quality » of the sampling frame
- The inactive units raise some difficulties : for those included in the sample, it is difficult to obtain responses during the observation phase, and it leads to a less effective sample size, as to difficulties when dealing with non responses

3.7. Sample design (2)

- Example (SIRUS): within the statistical registers, units (for example legal units) may be economically inactive, even if not closed from a legal point of view (in SIRENE)
- A variable « statistically active / inactive » has been introduced (mainly using tax information)
- This variable may be used when building frames, or, if the « presumed inactive units » are still included in the sample of a statistical survey, when dealing with non responses
- This information is shared by all business statisticians

3.7. Sample design (3)

- *Taking into account the statistical burden* : information about the burden (total burden, distribution among categories of businesses) is a recurrent demand
- For this purpose, a question is asked in every statistical survey about the time to fill the questionnaire
- The surveys for which a given unit has been sampled are also recorded in the statistical business register
- This information is also used to make sample coordination, in order to spread, as much as possible, the burden - *methodological reference* : *Sampling coordination of business surveys: a new method implemented at INSEE, Emmanuel Gros, on Eurostat website*

4. Design collection

4. Design collection (1)

- The data collection may use directly the statistical unit, but it also may use data from other units
- This is particularly the case when using administrative data, available at the level of legal units
- In this case some « micro-data » translator may be needed
- One may also distinguish the observation unit, and the reporting unit (even if, in most cases, they are the same)

4. Design collection (2)

- Example : for the French surveys using the enterprise as statistical unit :
 - For the largest groups, collection of data at the level of the enterprise
 - For the other ones, the collection unit used for the statistical surveys remains the legal unit → we draw cluster samples (of legal units belonging to the sampled enterprises), and the links existing, within the statistical register, between each enterprise and the legal units are then essential

Questions ?

Philippe.brion55@gmail.com