

# Application of a revised classification in the statistical programm: sampling and weight issues

Valentin Todorov<sup>1</sup>

<sup>1</sup>United Nations Industrial Development Organization (UNIDO)  
v.todorov@unido.org

Regional Workshop on ISIC Rev.4 implementation

Amman 28.09-1.10.2014



# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification
- 3 Weighting under the new ISIC classification
- 4 Summary and conclusions
- 5 References



# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification
- 3 Weighting under the new ISIC classification
- 4 Summary and conclusions
- 5 References



# Implementing a national version of ISIC Rev.4

## Overall implementation process

- Implementing a classification of economic activities, ICIC Rev 4, is a major task for all all countries. A substantial amount of work will need to be carried out in the fields of business registers, business surveys and national accounts.
- The overall implementation process can be separated into three different components which are closely related and will depend on each other. These elements will be discussed in detail, now follows an overview.

The components of the implementation process are:

- ① Structuring national classifications
- ② Adaptation of business register
- ③ **Sample designs and weights for surveys**

# Application in statistical programs

## Challenges faced in sampling and estimation

- Sampling procedures have to be updated for the new classification
- Scope of categories may have changed (new economic activities in scope, etc.)
- Should (at least for the transition period) produce estimates for old and new classification
- Overall sample size may increase
  - To increase reliability of dual-coded results
  - To compensate for startup problems of the new classification (misclassification of units)
  - To allow for increase of detail required in new classification
  - To better evaluate coherence between old and new classification



# Application in statistical programs

- The amount of work regarding sampling and estimation depends on the level of detail for back casting
- The change of industrial classification provides an excellent opportunity to substantially improve sample and estimation design
- A major drawback of such redesigns is that it often has systematic effects on the outcomes of the survey, leading to discontinuities in the series.
- An important aspect of a survey redesign is to minimize this inconvenience for data users. This can be accomplished by quantifying the effect of the redesign on the outcomes of the main parameters.



# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification**
- 3 Weighting under the new ISIC classification
- 4 Summary and conclusions
- 5 References



# Sampling under new ISIC classification

All business surveys are currently operated under ISIC Revision 3: the sample normally is selected with the industrial classification as one of the stratification variables

⇒ it will be necessary to redesign these surveys so that ISIC Rev. 4 is the industrial classification by which samples are selected.

## Types of sampling schemes

- There are different types of sampling schemes available.
  - cut off designs,
  - panel designs,
  - simple random sampling,
  - stratified simple random sampling,
  - systematic sampling.
- Probably the most common designs in use in National Statistics Institutes around the world are cut-off designs, stratified designs and combinations of the two.



# Impact on cut-off designs

## 1. A single cut-off threshold, applied to all industries

- Easy, since the cut-off thresholds do not differ by industry.
- The (single) threshold can remain the same even under the new coding structure.
- Production of back data is an issue - not a sampling problem.
- Take care of the overall sampling size (new industries included?)  
⇒ this may lead to a change in the threshold applied to all industries.
- Use **Register** information on counts in order to determine whether there is any need to amend the cutoff threshold.
- Also register information on relevant auxiliary variables (turnover, employment) may be used.



# Impact on cut-off designs

## 2. Different thresholds by industry

- More complicated, case since it is necessary to determine what the thresholds should be for the industries in the new industrial classification.
- Use **Register** information to analyze the numbers of businesses in each industry above certain thresholds.
- Combine this information with the correspondence tables to determine estimated thresholds for the new industries (initial thresholds).
- Care will need to be taken on the non 1-to-1 types of correspondence.
- Iterative approach - the counts of businesses and sums of relevant auxiliary variable in each industry will differ as the quality of that information on the business register improves.



# Impact on panel designs

## Impact on panel designs

- **Panel design:** where the same businesses are in the sample each period. No estimation of population totals; rather search for the change in the variable being measured, from one period to the next.
- Often used in short-term statistics to measure change; results derived from the panel are applied to a benchmarked total from a more reliable annual survey and are revised every year or so when the latest annual data become available.
- Much of the approach for cutoff designs can be applied.
- Use the opportunity to refresh and update the panel - essential to ensure good coverage is obtained across the range of industries in the new classification.
- Modeling and benchmarking techniques may need to be used to estimate missing totals for new industries or new size bands.

# Simple random sampling

## Simple random sampling

- The industrial classification may not feature in the sample design, but is used for tabulation by industry
- Very good results, when the sample size is very large.
- Easy to deal with - all the work is in the post-stratification of the businesses which have responded.
- There is an issue in terms of producing back series, if this is required.
- Simple random sampling with selection according to the industrial classification (but no stratification by size) is a stratified design - to be considered later.



# Stratified random sampling

## Stratified random sampling

- Stratified random sampling is the usual choice of operation of business surveys.
- The level of detail in ISIC can vary
- Allocation of the total sample size to strata is usually done by the **Neyman Optimal Allocation method** (Neyman 1934) - the sample size  $n_h$  and stratum  $h$  is:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$



# Stratified random sampling

## Stratified random sampling (cont.)

- After the reclassification, we can determine the population size in each of the new strata. Since  $S_h$  relates to the value in the population at large, we usually estimate this by  $\hat{S}_h$ , the variance of elements in stratum  $h$  in the sample.
- However, under newly-defined strata, we may not have these for some strata → alternative approaches needed.
- If some suitable auxiliary information (eg turnover, employment) is available, x-optimal allocation (Sarndal et al 1992) could be considered:

$$n_h = n \frac{N_h S_{xh}}{\sum_{h=1}^L N_h S_{xh}}$$



# Reallocating the sample in practice

## Top up samples

- An iterative approach: analyze with new information when it becomes available.
- Conduct analysis of numbers of businesses in old and new codes on the register and in survey samples.
- See if any look particularly small and try to re-allocate some of the sample to these, or increase the sample size - timing is important here.
- If publishing on both old and new codes simultaneously → necessary to boost the total sample size, to ensure an adequate sample for both. Later the sample could be cut back to its original size.
- Once enough information to estimate the variances under the new codes is available → a complete re-allocation.
- Take care when weighting the sample to correctly represent the sample design.

# Sampling under new ISIC classification

## Example for a short-term survey: time table for transition

Year $t_0$ , January	BR recoded to old and new classification
Year $t_0$	Continue sampling according to old NSIC but use BR information to tabulate the existing sample against the new strata in new NSIC. Use this scheme as the first attempt at the sample on the new basis. For strata that are weak, in terms of their sample size, estimate the numbers required for acceptable results on the new NSIC and seek approval to carry out this top-up of the sample
Year $t_0 + 1$ , January	Draw the sample on new NSIC (initially: old sample tabulated against the new industries and new strata, with weakest strata under the new NSIC boosted by a top-up of the sample)
Year $t_0 + 1$	Calculate variance of the elements in the sample in each of the strata based on returned data. Use sample variances as approximations of the population variances and use Neyman allocation to calculate new samples.
Year $t_0 + 2$ , January	Re-allocate the sample based on reported data (in a more efficient way now). The top-up sample can cease. Survey can be conducted according to new NSIC.





# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification
- 3 Weighting under the new ISIC classification**
- 4 Summary and conclusions
- 5 References



# Weighting under the new ISIC classification

## Weighting under the new ISIC

- Weighting in the context of estimation of variables using sample surveys, not in terms of weighting indices, such as occurs in the domain of short term statistics.
- Calibration estimation: finding weights which will modify the usual estimator such that the calibrated estimator for the auxiliary variable is one for which there is no calibration error.



# Weighting under the new ISIC classification

Three options for calibration of weights (listed in increasing order of risk and benefit):

<b>Option 1</b>	
Year 1	(i) Calculate calibration factors using the old classification; (ii) Produce results using conventional estimation for the old classification and by domain estimation for the new classification.
Year > 1	(i) Calculate calibration factors using the new classification; (ii) Produce results using conventional estimation for the new classification.
<b>Option 2</b>	
Year $\geq$ 1	(i) Calculate calibration factors using the new classification; (ii) Produce results using conventional estimation for the new classification and by domain estimation for the old classification; (iii) Variances for the old classification domains would need to be calculated differently (domain estimates) to those under the new system.



# Weighting under the new ISIC classification

<b>Option 3</b>	
Year 1	Calculate calibration factors using both classification systems; (i) In this case the population totals are reproduced by summing the weighted employment (turnover) for both classifications; (iii) Note that the variances for both classifications would need to be calculated using Statistics Canada's Generalized Estimation System (GES) (Estevao et al 1995).
Year > 1	(i) Calculate calibration factors using the new classification only; (ii) Produce results using conventional estimation.

- **Option 1** Discontinuity in the year after the change; maximum time is allowed for the new classification to settle down
- **Option 2** Discontinuity coincides with the year of change; risk due to using the new classification on the a year earlier than in option 1
- **Option 3** Risk that some unexpected weights are produced in year 1



# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification
- 3 Weighting under the new ISIC classification
- 4 Summary and conclusions**
- 5 References



# Summary and conclusions

- The outset that the change of classification requires extra resources for **sampling** and **estimation**.
- **Sampling under new ISIC classification:** The impact of the change on different types of surveys was considered: cut off designs, the use of panels, simple random sampling, stratified simple random sampling; the most common designs in use in National Statistics Institutes are cut-off designs, stratified designs and combinations of the two.
- **Weighting under new ISIC classification:** Three options for applying calibration weighting in the context of the classification change.



# Outline

- 1 Introduction
- 2 Sampling under new ISIC classification
- 3 Weighting under the new ISIC classification
- 4 Summary and conclusions
- 5 References



# References I



UNSD, EUROSTAT, UNIDO, Statistics Austria.

Different presentations and other materials from seminars and workshops.



EUROSTAT-3

Methodological aspects related to sampling designs and weights estimations





# Implementing a national version of ISIC Rev.4

**THANK YOU FOR YOUR ATTENTION**

`v.todorov@unido.org`

