Development of new statistical methodologies and indicators in selected areas of statistics in line with EU statistical standards

Ukraine

MISSION REPORT

on

Sample selection for the Capital Investment Survey, continued

Component no 3.1

Mission carried out by Tiina Orusild, Statistics Sweden

11.03.13-15.03.13

Version: Final

IPA 2007

Tiina Orusild
Statistics Sweden
Box 24300
SE –104 51, Stockholm
Sweden
Tel: +46 8 506 942 03
Email: tiina.orusild@scb.se

Table of contents

List of Abbreviations

ToR   Terms of Reference

SSSU   State Statistics Service of Ukraine

1. General comments

This mission report was prepared within the Twinning Project „Development of new statistical methodologies and indicators in selected areas of statistics in line with EU statistical standards". It was the fourth mission to be devoted to discussions on the sampling frame construction and sampling in the Capital Investment Survey within Component 3.1 of the project. The mission was aimed at defining a strategic plan forming the base of the further implementation of the project in this statistical area.

The concrete objectives of the mission were:

- To discuss sampling frame construction and sample selection

- To give recommendations on non-response adjustment

- To give recommendations on quality reporting

The consultants would like to express her thanks to all officials and individuals met for the kind support and valuable information which she received during the stay in Ukraine, and which highly facilitated the work of the consultant.

This views and observations stated in this report are those of the consultant and do not necessarily correspond to the views of EU, SSSU or Statistics Sweden.

2. Assessment and results

2.1 Introduction

The Quarterly Capital Investment Survey 2013 will be conducted with the current methodology using cut-off sampling with total enumeration of large and medium sized enterprises and including only some of the small enterprises. The small that are included have invested according to previous runs of the survey or are known to invest according to information from the regional offices.

In addition to the current methodology survey a pilot survey will be conducted including a totally enumerated part, large and medium sized enterprises and "atypical" enterprises, and probability sampling of small enterprises.

The two samples differ; the current method and the pilot survey, in that in the latter a probability sample of small enterprises will be selected while only a few small enterprises will be included in the current survey, i.e. the coverage of small enterprises will be improved in the pilot survey. The coverage of small enterprises in the cut-off survey used today is not satisfactory.

The results from the pilot survey using probability sampling will be assessed and the idea is to implement probability sampling in the Quarterly Capital Investment Survey in 2015.

2.2 Sample design

The 2013 quarterly pilot survey will be conducted as follows. All large and medium sized enterprises will be included with probability one. Among small enterprises "atypical", or outliers, are detected using the 3σ-method, see below, using data from the Annual survey 2011. The enterprises identified with this method are also included in the totally enumerated part. In addition some outliers are identified manually and also these are totally enumerated. An enterprise is defined as large, medium sized or small according to the following;

    a) Small enterprise: number of employees ≤ 50 and turnover ≤ 70 million hryvnia

    b) Medium: 50 < employees ≤ 250 and/or 70 million < turnover ≤ 100 million hryvnia

    c) Large: more than 250 employees > 250 and turnover > 100 million hryvnia

A sample of small enterprises is selected as follows. The small enterprises, excluding outliers, are stratified by economic activity, institutional sector and size. The stratification by economic activity is defined by two-digit NACE, institutional sector is divided in two groups' non-financial enterprises and other sectors. The size stratification is by number of employees and the following size classes are used

1. 0 employees,
2. 1-2 employees,
3. 3-5 employees,
4. 6-9 employees,
5. 10-19 employees and
6. 20-50 employees.

Institutional sector or economic activity is incorrect for some enterprises leading to some "additional" strata with only a few enterprises. New enterprises are stratified only by economic activity since there is no information in the Business register on size or institutional sector for them.

One of the following three different "strategies" is used in each economic activity (defined by 2-digit NACE):

1. Select all enterprises (total enumeration),
2. Select all large, medium sized and outliers,
3. Select all large, medium sized and outliers and select a stratified sample of small enterprises.

In Strategy 3, simple random sampling is used within strata. Strategy 1 is used in economic activities with few enterprises. Strategy 2 is used in some economic activities where the capital investments in the totally enumerated part exceed 90 percent of the capital investments in the economic activity. In Strategy 2 cut-off sampling is used where all small (according to the definition in a) above) enterprises are excluded. Data from the latest available annual survey is used to decide whether the totally enumerated part covers more than 90 percent or not and outliers are moved to the totally enumerated part before calculating the share of investments in the totally enumerated part. For the quarterly survey year *t*, data from the annual survey year *t-2* is used. If the number of enterprises is not too small and the amount of capital investments in the totally enumerated part is not large enough to use Strategy 2 then Strategy 3 is used, i.e. a probability sample of small enterprises is selected.

As mentioned above new enterprises are stratified by economic activity only and a simple random sample is selected in each strata.

The method for detection of outliers, or "atypical" enterprises, was changed. The method that had been used was the following. An enterprise, *k*, in strata *h* was identified as an outlier as follows

$$\text{if } (x_k - \bar{x}_h) \geq 3 \cdot s_h \text{ then } x_k \text{ is an outlier} \tag{1}$$

where x_k is capital investments for enterprise *k* in the annual survey 2011, $\bar{x}_h$ is the mean value of *x* in stratum *h* and s_h is the standard deviation for the *x*-values in stratum *h*. The condition in (1) was used in each strata and up to three times, namely if the coefficient of variation (cv_h) in stratum *h* exceeded 2, i.e. $cv_h = \dfrac{s_h}{\bar{x}_h} > 2$ after the first use of (1) and removal of outliers then (1) was checked again, removing new outliers etc. With this method about 12000 enterprises were identified as outliers. Looking closer on the values of these outlier enterprises, many of them were by no means outliers. Since there are many enterprises with value 0 investments the outlier detection method identified a large part of those with values > 0 as outliers even if the values were really small. This was changed to using (1) only once and removing small values from the list of "atypical" enterprises with the result of

less than 1500 enterprises. The number of "atypical" is still large and they are probably not real outliers.

In addition to the outlier detection method described above some enterprises are identified as outliers by experts (manually), however these are not numerous.

A sampling frame for the quarterly survey was constructed. The frame size was about 650 000 enterprises. SSSU has decided to use a sample size of at maximum 100 000 enterprises. The totally enumerated part consists of large and medium sized enterprises, strata with less than 10 enterprises and outliers. The totally enumerated part consists of about 45 000 enterprises. The remaining sample size, about 55 000, was used as sample size for the small and new enterprises. The sample size is very large. If costs have to be reduced or respondents' burden reduced it would be possible to do that.

The sample size for new enterprises is 10% of the enterprises in each stratum, which gives us a total sample size of about 2500 for the new enterprises. The remaining sample size is about 52 250.

Next we consider sample sizes for strata with small enterprises. In some economic activities the small enterprises were cut-off since the totally enumerated part of those economic activities cover a very large part of the total investments in the economic activity. That is the small enterprises are not sampled in some parts of the population. For the rest of the economic activities a sample of small enterprises is selected.

The methodology proposed for allocation of the total sample was a "double Neyman allocation". First, Neyman allocation was used to determine the sample size in the domains defined by economic activity (2-digit NACE). After that, Neyman allocation was used again to determine the sample size in each stratum within each domain. This allocation method was discussed during the previous mission and the recommendation was to change the allocation. This has not been done.

When allocating twice as described above, in the first allocation the domains are considered as strata. The allocation for domains will be optimal for estimating the population total, but the precision (or sample size) in each domain is not controlled.

An alternative approach was discussed during this and the previous mission and we calculated sample sizes with this alternative approach. In short the alternative approach was to decide on a precision in each domain and calculate the domain sample sizes. Then use Neyman allocation for allocation within each domain. In the calculations the domains were defined by economic activities (on 2-digit NACE) and each domain consisted of a number of strata (institutional sector and size). We considered the case when the precision in the domains was given by the coefficient of variation and we used several precisions which resulted in several sample sizes to select from. See the mission report from December 2012 for more details on the alternative approach.

A minimum sample size of 10 enterprises was used. The sample sizes from the calculations were studied and some of them was changed, e.g. avoiding too large weights and in strata were the proportion of enterprises with investments is small the sample size should not be too small.

In some strata there are a large number of small enterprises and only a few of them have invested according to previous surveys. In some examples the enterprises in the smallest size class contributed with very small values. In the "worst case" there were about 85000 enterprises in size class 1, 0 employees, and only a few of them had small investments according to the previous annual survey. The cut-off sampling in Strategy 2 could be changed, excluding more small, for instance enterprises with 0 employees in some economic activities, but not all small as defined by a) above.

2.3 Quality reporting

On the last meeting we discussed quality concepts and quality declarations. We considered the Swedish Quality concepts and as an example the quality declaration for the Swedish Quarterly Investment Survey.

3. Conclusions and recommendations

Change the allocation method. The double Newman allocation is difficult to explain and to have enough sample size in domains is not taken into account.

The outlier detection method does not work very well. Too many enterprises, with small investments, are identified as outliers. This should be studied more carefully and the outlier detection method should be changed.

The cut-off of small enterprises should be studied in more detail. The Strategy 2 with cut-off of all small enterprises, i.e. enterprises with number of employees $\leq$ 50 and turnover $\leq$ 70 million hryvnia could be changed cut-off of a smaller part of the small enterprises. In some economic activities a cut-off limit of the smallest enterprises could work very well, e.g. 0 employees could be considered. The cut-off does not have to be the same in all economic activities.

Regions are not used in the stratification so it is important to make sure that the sample sizes in the regions are large enough. This should be considered also for other domains, e.g. institutional sector on a more detailed level than that used in the stratification.

The stratification by economic activity is done by 2-digit Nace. In some economic activities it could be reasonable to stratify on a more detailed level. This should be discussed with the staff working with National Accounts. Even if all National Accounts needs on level of detail cannot be fulfilled this should be discussed.

The data from the pilot survey need to be thoroughly analyzed. Different problems in estimation including; non-response adjustment, how to treat outliers in the estimation, how to handle enterprises that change (mergers and splits) and how to adjust for over- and undercoverage and variance estimation should be discussed. The sampling design, cut-offs, sample sizes and stratification etc. should also be analyzed before redesigning the survey.

Annex 1. Terms of Reference

Twinning Project
" Development of Ukrainian Statistical Methodologies and Indicators in Selected Areas of Statistics in line with EU Statistical Standards "

Terms of Reference
for a short-term Mission to the State Statistics Service of Ukraine

Component 3.1 Sample Survey of Capital Investment

Activity 3.1.5 Assessment of the results of formation of sampling population

Background information

Statistics Denmark in partnership with Statistics Finland, Statistics Lithuania, Central Statistical Bureau of Latvia, Statistical Office of Slovak Republic, INE Spain - National Statistical Institute of Spain and Statistics Sweden, implements in Ukraine "Development of New Statistical Methodologies and Indicators in Selected Areas of Statistics in Line with EU Statistical Standards" Twinning Project. The State Statistics Service of Ukraine (State Statistics of Ukraine) is the Beneficiary of this Project).

This action is being implemented under Component *3.1 "Sample survey of capital investment". The purpose of this Component is to develop methodological support for sample survey of capital investment with regard to small-sized enterprises.*

This action will contribute to achieving the abovementioned objective and reference indicators specified in the contract, namely: *development of methodological support for sample survey of capital investment with regard to small-sized enterprise harmonized with the EU standards. .*

Purpose of the Mission

The prior purpose of the mission is: *Improve knowledge and practical skills of SSSU concerning EU experience on sample survey of capital investment.*

Expected Results

1. Assessment of the results of formation of sampling population based on the Methodology on formation of sampling population of active enterprises for capital investment surveys.
2. Receiving recommendations on methods and algorithms regarding no-answers; quality report.
3. Discussing the mission report.

Actions

The tentative schedule of the Mission is the following:

Date: 11.03.12
Date: 15.03.12

Tasks to be fulfilled by SSSU to facilitate the Mission

The Beneficiary will ensure the following:
Attendance of experts of division, which is engaged in national observation of capital investments and conducts its organizational and methodological support.

Consultant and Partner

The Mission will be conducted jointly with:
Tiina Orusild, Statistics Sweden

The partner from the country-beneficiary will be:
L. M. Ovdenko – Director of Production Statistics Department;
M. M. Sobko – Deputy Director of Production Statistics Department;
O.A. Muslinsky – Head of Division for Statistics of Capital Investments and Fixed Assets;
I.M. Kladchenko, L.O. Terletska – chief experts-economists of the same division;
S.A. Petrusenko. – Head of the Production Statistics Unit of the Main Interregional Department of Statistics;
O.V. Gonchar – Head of Division for Methodology of Production Statistics in Goods and Services of the Scientific and Technical Complex of Statistical Investigations, PhD in Economics.

Timing

The mission will be conducted within five days in Ukraine.

Report

The summary report on the results of the mission should be submitted not later than two weeks after the mission is completed.

Annex 2. Persons met

M. M. Sobko, Deputy Director of Production Statistics Department
O.A. Muslinsky, Head of Division for Statistics of Capital Investments and Fixed Assets
O. Moskalenko, Division for Statistics of Capital Investments and Fixed Assets
S.A. Petrusenko. Head of the Production Statistics Unit of the Main Interregional Department of Statistics
A. Tovchenko, Head of Unit of Mathematical Methods, Analysis and Processing of statistical information
O. Tkachenko, Unit of Mathematical Methods, Analysis and Processing of statistical information

RTA Team:
I. Bernstein , RTA
S. Taranova, Interpreter