

TWINNING CONTRACT

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

Ukraine



MISSION REPORT

on

Frame construction and sampling design for the Capital Investment Survey, continued

Component no 3.1

Activity 3.1.4

Mission carried out by Tiina Orusild and Linda Wiese, Statistics Sweden

10.12.12-13.12.12

Version: Final

		 STATISTICS DENMARK

IPA 2007

Author's name, address, e-mail (keep the relevant information)

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

*Linda Wiese
Statistics Sweden
Box 24300
SE -104 51, Stockholm
Sweden
Tel: +46 8 506 948 21
Email: linda.wiese@scb.se*

Table of contents

1. General comments.....	4
2. Assessment and results.....	4
3. Conclusions and recommendations.....	6
Annex 1. Terms of Reference.....	8
Annex 2. Persons met.....	10

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 third 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 the draft methodology on sampling frame construction and sample selection
- Assessment of results and giving recommendations on improvement of methodology

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

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

2. Assessment and results

This was the third mission on sampling frame construction and sampling design for the quarterly Capital Investment Survey. The annual Capital Investment Survey was not discussed during this mission since, at the moment, there are no plans on changing the methodology for the annual survey. The annual survey will continue to be conducted as a census.

Topics discussed during this mission were:

- Sample sizes
- Allocation
- Strategies used for selection of enterprises
- The 90 percent cut-off rule
- The software used by the methodologists for the investment survey
- The next step in the process

Oleksiy Muslinsnsky, Head of Division for Statistics of Capital Investments and Fixed Assets and Tanya Gorbach and Andrey Yaroshevich, Methodologists at the Scientific and Technical complex for Statistical Research, gave detailed presentations on the sampling frame construction and sampling methodology and, the production system for sampling in the Capital Investment Survey. They showed the software, written in R and SPSS, they will use for allocation and estimation. All steps in the software are well documented and easy to use.

In the Capital Investment Survey, estimates are produced for the whole population as well as for subpopulations such as regions (there are 27 regions in Ukraine) or economic activities defined by NACE, or by combining e.g. regions and economic activity. Domains defined by Institutional sector are also important. The sampling design will be stratified sampling with simple random sampling within strata. The stratification is done by nonfinancial and other enterprises, economic activity (2-digit NACE) and size of the enterprises. (To stratify by all domains of study would lead to too many small strata). Large and medium sized enterprise as well as outliers and enterprises that are identified by subject matter experts to be of great importance are placed in a totally enumerated stratum in each economic activity. Small enterprises are stratified by size classes defined by number of employees and samples of small enterprises are selected or they are excluded from the survey, see the strategies below. The classification of enterprises into large, medium sized and small is defined by turnover and number of employees (“size by law”).

The totally enumerated strata will include, except from large and medium sized enterprises, two types of “non typical” enterprises. The first type is enterprises with high investments identified by experts. These enterprises are identified by experts as having large investments and are moved from the sample part to the totally enumerated part before the sample selection. The second type is enterprises detected by the 3σ -method. According to the 3σ -method, enterprises with investments that differs three (or more) standard deviations from the mean investments in each strata (economic activity | size | sector) will be identified as outliers and moved from the sample part. The 3σ -method method is used up to three times in the following way. First outliers are detected by the 3σ -method and removed from the strata and then, the coefficient of variation, cv , is calculated for the remaining enterprises in each stratum and if the cv is larger than 2 then the 3σ -method is used again. If the cv is still larger than 2, after the two runs of the 3σ -method then the 3σ -method is used a third and final time.

For each economic activity (defined by 2-digit NACE) one of the following three strategies is used:

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

Strategy 1 is suggested to be used for economic activities with few enterprises. Strategy 2 is used when the capital investments in the totally enumerated part covers more than 90 percent of the capital investments in the economic activity. Data from the latest available annual survey is used to decide whether the totally enumerated part covers more than 90 percent of the investments or not. 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 less than or equal to 90 percent then strategy 3 is used for the economic activity.

The strategy for an economic activity could be changed over time, i.e. a cut-off of small enterprises is used one year and a sample of small enterprises (no cut cut-off) the next. This will cause problems when change is considered, since it might be difficult to investigate if change in level of investments is due to a change in investments or due to change in methodology.

The outlier detection and removing of outliers from strata from which samples are selected is done before the sampling strategy is chosen. This means that the 90 percent rule is applied after removal of outliers among the small enterprises. In the future this will be done by using data from the annual survey year $t-2$. Using old data there is a risk that small enterprises that should be included would instead be excluded.

The survey will also include a sample of newly established enterprises for which little information is available in the register. They are stratified separately and 10 percent of the new enterprises are selected to be included in a separate sample.

The methodology proposed for allocation of the total sample was a kind of “double Neyman allocation”. First, Neyman allocation is used to determine the sample size in the domains. After that, Neyman allocation is used again to determine the sample size in each stratum within each domain.

When allocating twice, 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. Decide on a precision in each domain and calculate the domain sample sizes. Then use Neyman allocation for allocation within each domain. The domains were defined by economic activities and each domain consisted of a number of strata. We considered the case when the precision in the domains was defined as the coefficient of variation.

When designing a new survey and developing a new sampling and allocation methodology, it is usually preferable that the same persons stay in the project during the whole process. Unfortunately the methodologists will leave the project by the end of this year. Although most steps in the development of the sampling design and estimation will be in place, methodologist will be needed also next year for analysing the results of the survey and further development of sampling and estimation in the survey.

4. Conclusions and recommendations

5.

First outliers are identified by using old data on investments from the annual survey and moved from the sample part of the frame to the totally enumerated part. Then one of the three strategies for sampling is chosen. If the totally enumerated part covers more than 90 percent of the investments then a cut-off of small enterprises is used, i.e. strategy 2. If all big investors are removed from the small enterprises, the small enterprises will probably not cover 10 percent of the investments and will therefore be cut-off. There is no evidence that the set of outliers will be the same over time and to introduce the 90 percent-rule could lead to large undercoverage.

According to the 90 percent-rule the strategy for an economic activity could be changed over time. This will cause problems when calculating change. A small increase in investment for a large enterprise can result in a change of methodology, from selecting a sample of small enterprises one year to excluding all small enterprises the next year. When excluding up to 10 percent of the investments, the result would probably show a decrease of investments, no matter if the investments in the domain have increased or decreased. Changing the strategy too often in a domain is not advisable. The real change will be difficult to estimate and it will be difficult to explain the change to users.

A better solution would be to use a cut-off in economic activities for which we know that the smallest enterprises are contributing with a value close to zero. The cut-off part should not necessary be all small enterprises but e.g. a part of the small. For instance in one economic activity there were a large number of enterprises, more than 100 000, in the smallest size class (enterprises with 0 employees) but in previous years data only a few of them had made investments and none with large amounts. The contribution to the domain total was small and there was no evidence that this part was of any importance. The investments for small enterprises may be important in some domains, whereas small enterprises in other domains don't invest a lot.

The identification of outliers repeatedly using the 3σ -method should be further analysed at least when data is available from the survey. The outlier identification is done on strata level and some of the values are probably not very large compared to other values in the domain. Some outlier method needs to be considered also in the estimation.

Instead of using the double Neyman allocation, we recommend the approach mentioned in Section 2. Decide on a precision in important domains, calculate sample sizes and use Neyman allocation. One example of this approach is given in Appendix 2 in the report¹.

The results of sample size calculations should be checked. The number of strata is large but one could e.g. check those with small and large sampling fractions. The, from the calculations, resulting sample sizes are not always reasonable and changes should be made. For instance when we calculated the sample size in the economic activity with a large number of enterprises with 0 employees the resulting sample size was large due to that the number of enterprises was large. In that case we should change the sample size to something much smaller. Different precisions can also be used to in different subpopulations if some subpopulations are more important than others.

¹ Statistiska centralbyrån (2003). *SAMU: The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden*. Background facts on Economic Statistics, 2003:3.

Methodologists should be involved in the work on re-designing the quarterly Capital Investment Survey.

Annex 1. Terms of Reference



This project is funded by the European Union



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 3.1.4 Analysis of existing sets of data for sampling survey on capital investment of small enterprises.

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. Discussing the Draft Methodology on formation of sampling population of active enterprises for capital investment surveys and technical background for formation of sampling population based on the Methodology.

2. Approbation of the Methodology and technical background using real sets of data on capital investment.
3. Assessment of achieved results and getting recommendations from the experts on their further improvement.

Actions

The tentative schedule of the Mission is the following:

Date: 10.12.12

Date: 13.12.12

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

Linda Wiese, 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 four 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

S.A. Petrusenko. Head of the Production Statistics Unit of the Main Interregional Department of Statistics

L.O. Terletska chief expert-economist at the Division for Statistics of Capital Investments and Fixed Assets

O. Moskalenko

Scientific and Technical complex for Statistical Research

T. Gorbach, Methodologist

A. Yaroshevich, Methodologist

RTA Team:

I. Bernstein , RTA

S. Taranova, Interpreter