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

Component no 3.1

Activity 3.1.2 Analysis of real data on capital investment

Mission carried out by Tiina Orusild, Statistics Sweden

18.09.12 – 21.09.12

Version: Final

		STATISTICS DENMARK

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

Executive Summary

If report-core text- exceeds 4 pages

Include information to Project Leaders and the RTA.
Main conclusions and highlights from findings.

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 second mission to be devoted to Analysis of real data sets on capital investment with regard to small-sized enterprises 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:

- Improve knowledge and practical skills of SSSU concerning frame construction and sampling design for the capital investments survey

The consultant 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 Statistics Sweden.

2. Assessment and results

Frame construction and sampling design and related issues for the quarterly- and the annual capital investment survey were discussed in more detail during this second mission.

The frame construction is done in the following way; In November 15 each year a new version of the Business Register, BR, is constructed. The BR contains active enterprises. To construct a frame for the capital investment survey all enterprises in the BR is used but some non active enterprises are added to the frame. Especially in the financial sector and the government sector there are enterprises that are investing but they are not included in the BR as active enterprises. Before adding the non-active enterprises they are checked against the Uniform State Register of Enterprises and Organisations of Ukraine, USREO. The BR alone cannot be used as the only source for frame construction for the capital investment survey. Other sources such as previous runs of the survey or other surveys are also necessary to improve coverage of the population.

The annual survey includes all questions from the quarterly survey and some additional questions. The large and medium sized enterprises are included in both the quarterly and the annual survey and are thus reporting on investments first for all quarters and then for the whole year. SSSU are studying the possibility to skip the annual survey and to move the questions which are not covered by the quarterly survey to other surveys. This would decrease the response burden for the enterprises, especially for all large and medium sized enterprises that are now included in both the quarterly and the annual survey.

The planned changes of methodology from cut-off survey of the small enterprises to probability sampling have implications on the results of the survey. The improved coverage of small enterprises should increase the estimates of totals and it also effects calculations of indices. The question on how to deal with this problem was raised. The National accounts need estimates which are comparable over time. In 2013 the quarterly survey will be conducted with both the old and the new methodology and one question was if this should be extended to a longer time period than one year. The times series are short, from 2010 and forward, so the planned one year of using the old and new methodology would probably be enough for these purposes. The new sampling design is planned to be implemented in 2015.

In the quarterly survey, enterprises have been removed or added to the survey during the year, i.e. the set of enterprises in one quarter has not been the same in each quarter during a calendar year. If

probability sampling is introduced then adding or deleting enterprises should be avoided if e.g. a sample is selected once a year and the same sample is used during all four quarters. In probability sampling weighting is done to calculate estimates of e.g. totals and to calculate weights strata sizes and sample sizes (or number of respondents) are used. If enterprises that are not in the sample are added to the survey and/or enterprises in the sample are removed bias is introduced and this can be difficult to adjust for in the estimation.

The work on a new sampling design for the quarterly capital investment survey was presented by Tanya Gorbach from the Scientific and Technical complex for Statistical Research. The suggested design for the quarterly survey was the following. The sampling frame is stratified by size and economic activity. The stratification by economic activity was done by 2-digit NACE and the size of the enterprises is defined as follows;

- Small: employees $\leq$ 50 and turnover $\leq$ 70 million hryvnia
- Medium: 50 < employees $\leq$ 250 and/or 70 < turnover $\leq$ 100 million hryvnia
- Large: more than 250 employees > 250 and turnover > 100 million hryvnia

The large and medium sized enterprises are totally enumerated and the small enterprises are further stratified by number of employees using size classes 0, 1-2, 3-5, 6-9, 10-19 and 20-50. Among the small enterprises strata with less than or equal to 10 enterprises are totally enumerated and, outliers are identified and moved to a stratum which is totally enumerated. Among the remaining small enterprises, excluding small strata and outliers, a probability sample is selected or the small enterprises are excluded, see below.

The analysis was done using data on investments from one annual survey. The identification of outliers was done using the "3σ-rule" on these data. (The 3σ-rule for outlier detection means to look for observations that deviate more than three times the standard deviation from the mean).

A total sample size of maximum 100000 enterprises, including the totally enumerated part, was used. In the sample part Neyman allocation was used on the whole population of small enterprises that are not totally enumerated. Using Neyman allocation in practice was also discussed, e.g. what variables could be used and the resulting sample sizes.

An analysis of investments for the large, medium and small sized enterprises was done in each 2-digit NACE group. The analysis resulted in different strategies for different economic activities, e.g. if a large amount of the investments in an economic activity are covered by the large and medium sized enterprises then a cut-off can be used, i.e. the small enterprises are not included in the survey. If e.g. the large and medium sized enterprises do not cover a large enough amount of the investments in an economic activity then a sample of small enterprises would be would be selected in that economic activity.

Another problem that was briefly discussed was how to deal with splits and mergers of enterprises in the estimation. The methodologists have to consider this problem and set up rules for how this should be handled in practice. This problem will be further discussed during the study visit and on the next mission.

3. Conclusions and recommendations

Only one year of data has been used in the work on developing a sampling design for the quarterly survey. The cut-off of small enterprises, i.e. enterprises with less than or equal to 50 employees and turnover less than 70 million hrynas, in some economic activities should be further studied since only one year of data has been analyzed. The quality of the size measure in the BR is maybe not yet good enough and there could be small enterprises, according to the BR, which actually are important to the survey. The outliers have been removed from the sampling part i the analysis but if the suggested design is implemented it is not possible to find all outliers before sample selection.

The investments surveys, quarterly and annual, are the only source of information for investments i Ukraine. If only the quarterly survey would be conducted then this survey would be the only source of information and it would be important to be able to produce good quality estimates of totals and change from this survey. The introduction of a cut-off in some economic activities should be carefully studied. The idea to sometimes include and sometimes remove the small enterprises in the sample would lead to too many redesigns of the survey. It is usually better to construct a design that is more stable over time, cf. National Accounts and other users' needs.

The outlier detection method, the "3σ-method", is not the best method to use in this case. Consider using some other more robust method.

The sampling design that was presented during this mission was stratified sampling with simple random sampling within strata, SRS. The stratification was done by NACE, 2-digit level, and size (by law) where the small enterprises are stratified by size classes defined by number of employees. Important domains of study of this survey are economic activities and regions. Stratification by domains should be further studied. Regions or some grouping of regions should be considered and also the level of detail of the stratification by NACE. Maybe some economic activities should be stratified on a more detailed level and some on less detailed level.
Consider also stratification by size of the totally enumerated part. This could be used for nonresponse follow-up and for nonresponse adjustment.

When deciding on a sample size, important domains have not been considered. It is a bit unclear how the total sample size of 100000 enterprises has been calculated or if the size was only decided from costs. In many cases a sample size is decided by first determining a precision in terms of variance, length of confidence interval or coefficient of variation and from that try to calculate a sample size. This should be done for important (non-overlapping) domains that cover the whole population, so that the sample size is large enough in important sub-populations. When a sample size in a domain is determined then one can use e.g. Neyman allocation in the domain (if domains consist of strata). An example of this is that the domains are the economic activities on e.g. 2-digit level and these are stratified by size. In the suggested sample allocation Neyman allocation has been used for the whole sampling part without taking into account the sample sizes for domains. Sample sizes and Neyman allocation will be discussed during the study visit to Statistics Sweden.

The total sample size of 100000 is large. Precision requirements should be considered when deciding the sample size, se above. One question is would it be possible to select samples among e.g. the smallest medium sized enterprises instead of selecting all, reducing the total sample size?

Neyman allocation was used but some of the resulting sample sizes were very large or small. When using Neyman allocation the resulting sample sizes should be checked and unreasonable values can be changed before sample selection.

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 3.1.2 Analysis of real data sets on capital investment with regard to small-sized 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 enterprises.*

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

Getting familiarized with information contained in the Business Register. Identifying criteria to comply sample frame using the Business Register, Building algorithms on the

sample frame generation based on the previous years' observations. Estimation of the results obtained.

Actions

The tentative schedule of the Mission is the following:

Date: 18.09.2012
Date: 21.09.2012

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 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

SSSU:
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, chief expert-economist at the Division for Statistics of Capital Investments and Fixed Assets
L.O. Terletska chief expert-economist at the Division for Statistics of Capital Investments and Fixed Assets
S.A. Petrusenko. Head of the Production Statistics Unit of the Main Interregional

Scientific and Technical complex for Statistical Research
T. Gorbach

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