TWINNING CONTRACT

JO 21 ENI ST 01 22

Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices

MISSION REPORT

on

Component 3
Strategy for missing or erroneous data

Activity 3.1.1
Review methodology for treatment of missing values: Business statistics

Mission carried out by
Mr. Peter Tibert Stoltze, Statistics Denmark
Dr. Kai Lorentz, DESTATIS

Amman, Jordan
January 09-12, 2023

Version: Final

دائرة الإحصاءات العامة
Department of Statistics

STATISTICS DENMARK

DESTATIS Statistisches Bundesamt

Istat

Statistics Lithuania

Statistics Finland

Authors' names, addresses, e-mails

Mr. Peter Tibert Stoltze
Head of division
Methodology and Analysis
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3827
Email: psl@dst.dk

Dr. Kai Lorentz
Head of section
Mathematical-Statistical Methods for Household Surveys and Census
Statistisches Bundesamt (DESTATIS)
Gustav Stresemann Ring 11
DE-65189 Wiesbaden
Germany
Tel: +49 611 75 2589
Email: kai-lorentz@destatis.de

Table of contents

List of Abbreviations

- BC – Beneficiary Country
- CoP – General Population and Housing Census 2015
- DoS – Department of Statistics
- HEIS – Household Expenditure and Income Survey
- MS – Member State
- PL – Project Leader
- RTA – Resident Twinning Advisor
- SBR – Statistical Business Register
- STE – Short-term Expert
- ToR – Terms of Reference

- N – Size of population (number of units)
- n – Size of sample (number of selected units)
- m – Size of response (number of responding units)
- π – Inclusion probability (Greek letter pi)
- SRS – Simple random sampling
- STSI – Stratified simple random sampling

1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices". The activity reflected in the report was part of component 3.

According to the Terms of Reference, the purpose of this activity was to review current methodology for dealing with missing or erroneous data for surveys in the economic sector.

The consultants would like to express their sincere thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Jordan and which highly facilitated their work. The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, DESTATIS, or Statistics Denmark.

2. Assessment and results

The purpose of this mission (as described in the ToR, see excerpt in Annex 1) was to "review current methodology for dealing with missing or erroneous data for surveys in the economic sector." That was exactly what was done during the mission albeit in a slightly different form from what was prescribed in the same ToR: "This will take outset in BC and MS expert jointly audit the production process for the Economic surveys carried out by DoS and make recommendations."

The initial plan was to audit eight surveys (listed in Annex 1) within the business domain at DoS. During the first day it became clear, that it would be beneficial to devote some time to cover the *Comprehensive survey of establishments* from 2018, since this is the common frame for 6 of the 8 samples considered. As this frame is by now to some extend outdated, this seems to be a major source of (unit-) non-response for the economic sample surveys. Also, new-born establishments have not been included in the frame, so there is also an under-coverage problem. Therefore, of special interest was the plan for updating the information from this survey within the next two years according to the following plan:

- In 2023 statistical blocks with more than 35 establishments or at least one establishment with revenue above a certain threshold will be visited;
- In 2024 statistical blocks with 15-35 establishments will be visited.

The information from this comprehensive survey and the subsequent updating might be a valuable input to the planned statistical business register (SBR).

An audit of the Service Survey was initiated based on the methodological description and explanation from the domain experts. The key figures for the survey were as follows (using standard notation:

- Size of population: N = 47,000
- Size of sample: n = 3420
- Response rate (m/n): Approximately 75 percent
- Reason for non-response:
    - Establishments have closed down
    - Procrastination (delaying or postponing participation) and refusal (15 pct.)

- Change of economic activity (2-3 pct.)

The provided information was of quite high quality, and it was discussed that a thorough non-response analysis should be based on this type of information (with specific rather than approximate figures) for all surveys and preferably in more detail (e.g. per stratum) for the last 5 years (as per ToR).

Following the discussion of the Service Survey, a number of questions were put forward by the DoS staff revolving about the applicability of so-called *shadow samples* but also in relation to information from the sample which during the sampling process is discovered to have changed. An example is the following question:

A sampled (and responding) unit from a stratum with size class "10-19 employees" actually has 8 employees. What should be done?

The correct answer depends somewhat on context, but the following general advice applies:

- Do not change the initial inclusion probability π
- If estimating survey results for size classes, the unit in question should not be in the size class "10-19 employees" but rather "5-9 employees" with appropriate weight (design weight is $1/\pi$)

It was also discovered, that apparently *item non-response* does not occur within the DoS surveys, only *unit non-response*. In other words, either a survey response ("questionnaire") is completed or not – there are no cases of partial response. This is a somewhat special situation, but since most (or all) interviews are carried out as CAPI (computer-assisted personal interview), this can mean that the interviewers are very careful in doing a good job. However, the absence of item non-response lessens the need for imputation techniques.

During the first day it also became clear that a general review of some terminology and methods for *stratified simple random sampling* (STSI) would be beneficial in order to align understanding of basic concepts between BC and MS experts. This exercise was carried out as an interactive workshop with active participation from the MS experts. Unfortunately, staff members from the methodological unit of DoS were not present.

Many important aspects were covered, but it is worthwhile to highlight a few of these:

- Keep the original frame as long as any sample from this frame is still in use
- Be aware of the concept of stratum as opposed to domains
    - Stratum is the group of enterprises (generally: primary sampling units) from which the sample is chosen
    - Domain is the defines group of units over which an estimate is calculated, e.g. a combination of region and activity class
- Prior to data collection is good practice to describe the sample by constructing a dataset on stratum level with information about stratum size (N_h), sample size in stratum (n_h), and design weight ($dw = N_h/n_h$)
- Following data collection, it is good practice to augment the above-mentioned data set by information about response on stratum level (m_h) and possible stratum-wise response rates ($R_h = m_h/n_h$)

There was general interest in imputation methodology, both about specific techniques but also regarding the application of imputation, i.e. when to do it. Since item non-response is not present, imputation is only relevant for unit non-response. However, for small and medium sized enterprises imputation is not recommended. Instead the recommended practice is adjustment of weights for non-response.

For large enterprises, imputation may be a possibility since adjustment of weights for the strata with the largest units (which generally have design weight equal to 1) should be avoided. Imputation methods falls into two major groups: Donor imputation and predictive modelling. General principles were covered albeit not in detail. The application of specific techniques is generally dependent on the available software and could be a topic for subsequent missions. However, one should remember that even if we apply a statistical model to do imputation, we are basically guessing at numbers, that we were not able to secure in the preferred manner. Another caveat of imputation is the added uncertainty, which is technically quite difficult to account for. If imputation is applied, it should be thoroughly described in the documentation of statistics.

Quite a lot of time was devoted to the discussion of the concept of *objective* and *subjective* non-response which is crucial in order to make correct adjustment for non-response to the design weights avoiding bias. During the initial audit of the Service Survey it was demonstrated that unit non-response is attributed to one of eight different reason, e.g.

- Establishments have closed down
- Establishment has moved
- Establishment has new owner
- Procrastination (delaying or postponing participation) and refusal
- Change of economic activity
- Establishment is not main branch

These categories are very suitable as the starting point for non-response adjustment where the crucial question is: *Is zero the correct value for the variable in question?* A number of examples were constructed and discussed.

3. Conclusions and recommendations

- The analysis of non-response – like worked out for the Service sector survey during the visit – should be carried out for all the sample surveys in the economic sector by DoS. And this should not be done only once, but on a regular basis.
- The planned update mechanism for the establishment frame should be carried out and new samples should then be drawn from the updated frame.
- Staff members from the methodological unit should assist the subject matter statisticians in these tasks and also participate in the methodological mission activities of the Twinning project.
- The knowledge of sampling methodology should be further strengthened within DoS. The methodological unit should be leading and assisting the subject matter statisticians in the correct application of sampling methodology.
- In view of a future business register, sampling plans for the business surveys should be revised. This includes the question of appropriate statistical units for the sample surveys and requires the participation of staff members from the Methodological unit and National Accounts for these tasks.
- Future missions on methodological issues should be carried out with emphasis on practical data-driven aspects. This means that discussions, explanations and

recommendations during the mission should be clarified using some real (or example) data sets from DoS. This is important in order to agree upon common terminology and avoid misunderstandings.

Annex 1. Terms of Reference (excerpt)

1. Purpose of the activity

The purpose of this activity is to review current methodology for dealing with missing or erroneous data for surveys in the economic sector. This will take outset in BC and MS expert jointly audit the production process for the Economic surveys carried out by DoS and make recommendations.

Audit of the production process of the Industry Survey and the Service Survey in DoS
The audit will be a professional jointly review by BC and MS expert of the production process of surveys in the economic sector in DoS with focus on process steps that can affect non-response or erroneous data. In order to systemize the audit it will follow the production steps in the Generic Statistical Business Process Model (GSBPM).

Non-response
During the mission, examination of pattern of non-response will be discussed.

Unit nonresponse refers to the complete absence of an interview from a sampled unit whereas item nonresponse refers to the absence of answers to specific questions in the interview after the sampled unit agrees to participate in the survey. In DoS only unit response does occur in economic surveys.

Altogether, eight economic surveys is carried out in DoS. They are:

- Industry Survey
- Service Survey
- Internal Trade Survey
- Transport Survey
- Construction Survey
- Finance Survey
- Insurance Survey
- Employee Benefits Survey

Methodological descriptions for the first four of the listed surveys are included as appendixes to this ToR.

In addition, quality of surveys will be discussed taking outset in [ESS guidelines for the implementation of quality and performance indicators (QPI's)]

2. Expected output of the activity

- A common terminology is introduced;
- Methodologies used in economic surveys in DoS will be scrutinized by BC and MS expert together with main focus on Missing a;
- Indicators for missing data have been calculated for selected economic surveys;

Annex 2: Programme for the mission (planned)

Date	Time	Venue	Topic
Monday 09 January 2023	Meeting room – 1st floor	09:00 – 09:30	Welcoming, acquaintance, program of the week (Mr. Mohammad AbelRazzaq and Charlotte Nielsen)
		09:30 – 10:00	**MS:** Introduction to the audit (Peter Stoltze)
		10:00 – 12:00	**BC and MS:** Audit of the Service Survey Moderator:
		12:00 – 13:00	Lunch break
		13:00 – 15:00	**BC and MS:** Audit of the Industry Survey Moderator:
Tuesday 10 January 2023	Meeting room – 1st floor	09:00 – 09:30	**MS:** Summary and conclusion from day 1
		09:30 - 11:30	**BC and MS:** Audit of the the Internal Trade Survey Moderator:
		11:30 – 12:30	Lunch break
		12:30 – 14:30	**BC and MS:** Audit of the transportation and construction survey Moderator:
Wednesday 11 January 2023	Meeting room – 1st floor	09:00 – 09:30	**MS:** Summary and conclusion from day 2
		09:30 – 13:00	**MS and BC:** Audit of the Finance Survey, Insurance Survey, Employee Benefits Survey Moderator:
		12:30 – 13:30	Lunch break
		13:30 – 15:00	**MS:** Introduction to Quality Performance Indicators (QPI's)
Thursday 12 January 2023	Meeting room – 1st floor	09:00 – 09:30	**MS:** Summary and conclusion from day 3
		09:30 – 12:30	**MS and BC:** Summarizing the results and next steps
		12:30 – 13:30	Lunch break
		13:30 – 14:00	**BC:** Presentation of results for the management
		14:00 – 14:30	**MS and BC:** Evaluation of the mission and final remarks and thanks

Annex 3. Persons met

DoS experts:

Muhammad AbdelRazzaq - Director of Economic surveys (DoS Component Leader)

Faten Al-Ramini - Head of Services, Internal Trade, Financial and Insurance Division
Safa' Hamed Ekreem - Services Division
Areej Al-Shea'ar - Services Division
Rasha AbdelKareem Abu-Roman - Services Division

Mazen Khalifeh - Head of Industry and Energy Division
Noor Al-Kharabsheh - Industry and Energy Division
Nisreen Alaween - Industry and Energy Division
Khaldoon Kena'an - Industry and Energy Division
Raba'a Al-Jabrah - Industry and Energy Division

Amal Al-Samawi - Head of Construction Division
Ashraf Al-Hajaj - Construction Division

Samaher Ali Al-Haj - Transport and Tourism Division
Mwafaq Abu-Matar - Transport and Tourism Division

Walid Battah - Director of National Accounts Directorate
Eman Assaf - National Accounts

Twinning team:

Eng. Mohammad Khalaf, Director of Sustainable Development Unit, Department of Statistics, Jordan (RTA Counterpart)
Ms. Charlotte Nielsen (RTA)
Ms. Zaina Amireh (Language Assistant)