TWINNING CONTRACT

AZ/16/ENI/ST/01/19 (AZ/53)

Support to the State Statistical Committee and the State Tax Service under the Ministry of Economy to strengthen collection, harmonization, analysis, publishing and dissemination of business statistics

MISSION REPORT

Activity 4.4.C

Activity Title: Processing SBS data

Component 2: The production of business statistics/ SME statistics in terms of content and organization is upgraded

Mission carried out by

Mr. Guus van de Burgt, Statistics Researcher, Statistics Netherlands

Dr. (Mr.) Andrius Čiginas, Chief Specialist, Statistics Lithuania

5-8 April 2022

On-site mission at the premises of the State Statistical Committee of the Republic of Azerbaijan

Version: Final

Expert contact information

Mr. Guus van de Burgt

Statistics Researcher, Statistics Netherlands

Email: ghj.vandeburgt@cbs.nl

Dr. (Mr.) Andrius Čiginas

Chief Specialist, Statistics Development Division, Statistics Lithuania

Email: andrius.ciginas@stat.gov.lt

Table of contents

List of Abbreviations

AD	Administrative Data
ASAN	State Agency for Public Services to Citizens of Azerbaijan
AZ	Azerbaijan
BC	Beneficiary Country
EBS	European Business Statistics
EC	European Commission
ESS	European Statistical System
EU	European Union
MLSP	Ministry of Labour and Social Policy
MS	EU Member State
NACE	Nomenclature générale des Activités économiques dans le Communautés européennes NACE is the industrial classification of units dealing in economic activities
NSI	National Statistical Institute
RS	Remote Sessions
RTA	Resident Twinning Adviser
SRSU	State Register of Statistical Units
SBS	Structural Business Statistics
SSC	State Statistical Committee
STS	Short Term Statistics (on turnover)
ToR	Terms of Reference

Twinning Project: AZ/16/ENI/ST/01/19 (AZ/53)
"Support to the State Statistical Committee and the State Tax Service under the Ministry of Economy to strengthen collection, harmonization, analysis, publishing and dissemination of business statistics"

1. General comments

This mission report was prepared within the EU Twinning Project "Support to the State Statistical Committee and the State Tax Service under the Ministry of Economy to strengthen collection, harmonization, analysis, publishing and dissemination of business statistics". It was the eighth mission to be devoted to Result 4 within Component 2 of the project.

The purposes of the mission were:

- R code for unit imputation of SBS data developed
- Recommendations for further development and testing imputation
- Possibilities to develop validation and editing identified
- Mission report

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

These views and observations stated in this report are those of the consultant and do not necessarily correspond to the views of EU, SSC, STS, SME-Agency, Statistics Finland, Statistics Lithuania, Statistics Netherlands, Statistics Denmark, and HAUS.

2. Status at the beginning of the mission

- Need for unit imputation was identified
- Structure of SBS-data was available
- R code for imputations based on this structure was available, but not yet tested since no access to data

3. Status of mission results

- R code for unit imputations for SBS is running on actual SBS and Short-term statistics data of SSC for 2020, also using 2019 data.
- The hierarchy of imputation models designed in the previous mission has been simplified for practical purposes.
- Quality of this code is tested by replacing actual data with imputations and looking at differences
- Participants have been instructed on how to use this R code.
- Progress in sessions and delivery of data was slower than expected. Participants had a strong preference for a tool they can immediately use instead of advice. Therefore, there has been little attention to validation and editing.

4. Sustainability of the achievements

Having an R code for unit imputation is step 1. Using it is step 2. That is the most important step. We feel confident that SSC will use this code. We are in strong doubt if it will actually be used to reduce sample size (for small units) or only as a tool to have total figures in an early stage when not all units have yet responded. It would be nice to stay in touch and see how this develops.

5. Recommendations for the future (short and long term)

Since this was the last mission in this twinning project, SSC will have to perform these recommendations by itself.

- Examine in the short term what more missing small units e.g. missing 50% would make for a difference in total figures. Based on this research in the medium-term a smaller sample is possible for small units and/or less effort put into observing each one. When making a smaller sample, the sample ratio may be different per

NACE group. Choose what variable to optimize in the sample design. Turnover is usually more important for business statistics but value added is for National Accounts. Optimizing on value-added is recommended.

- Data structure of SBS can be improved, so the meaning of a variable that respondents filled in can be used for both imputations and output. The current structure is variable, value, sub value, sub value, and so on. Where number and meaning of sub-values depend on the questionnaire and what the respondent indicated in the top row, e.g.

Section I İncome   **manat**

The name of indicators	row code	Total for the enterprise	by main activity	including:* for non main activities (by indicating the activity code)		
				code	code	code
B	1	A 2	3 B	4 C	5 d	
Income from the sale of goods, works and services (including returned goods and trade discounts, excluding VAT, excise tax) - total	100					

Section I İncome   **manat**

The name of indicators	row code	Total for the enterprise	by main activity	including:* for non main activities (by indicating the activity code)		
				code	code	code
B	1	2	3 C	4 D	5 E	
Income from the sale of goods, works and services (including returned goods and trade discounts, excluding VAT, excise tax) - total	100					

Column 4 of enterprise one has the same meaning as column 3 of enterprise 2, it is both activity C. But in the data we do not see this activity code and add column 3 with column 3 when imputing so we get some combination of activities B and C. That is hard to interpret. By making a data structure in this format:

Enterprise	Variable	Variable meaning	Sub variable	Value
1	100	Income from the...	Total	...
1	100	Income from the...	Activity A	...
1	100	Income from the...	Activity B	...

the meaning of the variables is available for both imputations and output. And imputation will be easier since you only need to impute 1 column (value) instead of 13 (d1-d13).

- In the code we made we read SBS and Short-term statistics (STS) data of many different excel-files. And the output was only made in R. It is recommended for SBS data to make a direct link to SQL to read these data and also to write the values of the imputations to that SQL database. Immediate reading from SQL is less work and you will always have actual data. By writing the imputations to SQL all needed SBS data stays in one place. Imputed units already have the same format as the observed units to make it easy to write them in SQL. An example code for reading SQL data is already provided. For Short-term statistics, data only needs to be read once a year, so there is less need for a direct link to SQL.

- Examine how to handle extreme developments in STS (>10 or <0.1) if they are influential on outcomes.

- Different boundaries (currently we have taken the requirement to have at least one unit) in model choice for e.g. structure of NACE5 digit.

- Add more of the presented imputation options if needed, e.g. use development of turnover in tax if no development of turnover in STS is available. Or scale level of imputation to level of turnover in STS if STS for current period is available but best imputation method could not be applied due to either missing SBS and / or STS of previous period.

- Once tax data is available you might use those instead of unit imputation to derive SBS-estimates for units. Advantages of using tax data is more variables are available and probably more units are available than in STS. When using tax data for SBS primary observations are only needed for the SBS-variables that are more detailed than tax variables. You might use ratios of previous year or of higher size class that is still primary observed to derive those more detailed variables. You will first need to examine where it is possible to replace SBS with tax data. Perhaps for certain NACE or for bigger units this is not well possible. But first results looked promising as was examined in an earlier mission.

- Apply imputations to other processes e.g. in Short-term statistics.

- If many people work with imputations for SBS then use the R-code to have IT build it into the current software for SBS.
- For the long term converting to an online IT infrastructure has benefits for missions like this and for regular work. If experts can work online with actual data in a secure environment or fake data that resemble the actual data, they can prepare more already in their own country in an online or hybrid mission, and therefore you will get even more results out of the missions. In addition at an onsite mission experts do not block someone's workspace. For regular work, it is easy to work at home or abroad if you have an online IT infrastructure. It is also easier to make different environments for developing, testing, and using the software.

6. Identification of needs for additional support

BC should be able to perform these recommendations by itself or apply for support if needed. Since this is the last mission in the project, we have no option to already offer support.

7. Outstanding issues

Tax data were supposed to become an important input for SBS. Since a continuous delivery of tax data from State Tax Service to SSC has been postponed, developing methods to transform tax data to SBS has been removed from this mission's agenda. Once SSC starts receiving these data it is recommended to use these data for (parts of) SBS.

Annex 1. Terms of Reference

> **Terms of Reference**
>
> **EU Twinning Project: AZ/16/ENI/ST/01/19 (AZ/53)**
>
> **Support to the State Statistical Committee and the State Tax Service under the Ministry of Economy to strengthen collection, harmonization, analysis, publishing and dissemination of business statistics**
>
> **Component 2: The production of business statistics/ SME statistics in terms of content and organization is upgraded**
> **5-8.4.2022 Hosting institution: SSC**

Activity 4.4.C: Processing SBS data

Result 4: Business Statistics in Azerbaijan redesigned
Indicator of Achievement:

4.1.4 The use of administrative information maximized to reduce response burden and costs associated with operational aspects of surveys.

Prerequisites: Activity 4.4.B

This activity will be carried out in a 4 Working Day's mission.

Purpose of the activity

Component 2 Result 4: Business Statistics (BS) in Azerbaijan redesigned

Expected Outputs of the Activity

- R code for unit imputation of SBS data developed
- Recommendations for further development and testing imputation
- Possibilities to develop validation and editing identified
- Mission report

Resources

State Statistical Committee (SSC)
- Mr. Nuru Suleymanov, Head of Dept., Dept. of National Accounts & Macroeconomic Indicators
- Ms. Sevinj Fattahova, Head of sector, Dept. of National Accounts & Macroeconomic Indicators
- Ms. Ilaha Babayeva, Chief Advisor, Dept. of National Accounts & Macroeconomic Indicators
- Mr. Turgut Shahmarli, Leading Adviser, National Accounts Statistics Department

Small and Medium-Sized Enterprises Development Agency
- Mr. Salman Humbatov, Leading Advisor, Secretariat

MS Experts
- Mr. Guus von de Burgt, *Statistics Researcher*, Statistics Netherlands
- Dr. (Mr.) Andrius Čiginas*, Chief Specialist, Statistics Development Division,* Statistics Lithuanian

Twinning Project Administration
- Mr. Ville-Matti Pilviö, RTA
- Ms. Gultakin Babayeva, RTAA

- Mr. Tarlan Arzumanov, Language Assistant

5-8.4.2022 - Agenda

Day/Time	Place	Event	Purpose / Details
Tuesday 10-17	SSC	Session 1	Assessment of the progress and identifying next steps in developing imputation **Unit imputation R-code** review
Wednesday 10-17	SSC	Session 2	Unit imputation R-code review Reviewing and validating data
Thursday 10-17	SSC	Session 3	**Validation** of indicators Editing data Summary
Friday 10-17	SSC	Session 4	**Identifying next steps** Main findings to an activity report