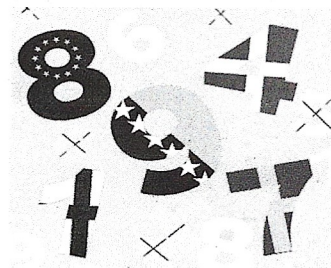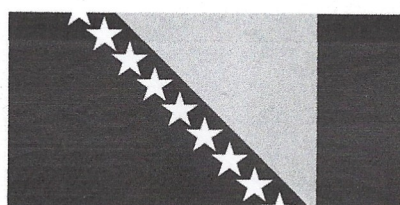# TWINNING CONTRACT

BA 17 IPA ST 01 20

Further Support to the Reform of Statistics System in Bosnia and Herzegovina

MISSION REPORT

Activity 2E1: Analysis of the Pilot Agricultural Census Results

Component 2 - Agriculture Statistics

Mission carried out by
Rudolf Seljak, Statistical Office of the Republic of Slovenia
19 – 22 December 2022

Version: Final

Expert contact information

Rudolf Seljak
Statistical Office of the Republic of Slovenia
Ljubljana, Slovenia
rudi.seljak@gmail.com

Table of contents

List of Abbreviations

BHAS	Agency for Statistics of Bosnia and Herzegovina
BiH	Bosnia and Herzegovina
CBBH	Central Bank of Bosnia and Herzegovina
EC	European Commission
EU	European Union
FBiH	Federation of Bosnia and Herzegovina
FIS	Institute for Statistics of Federation of Bosnia and Herzegovina
MS	EU Member State
RSIS	Institute for Statistics of Republika Srpska
RTA	Resident Twinning Adviser
ToR	Terms of Reference

1 General comments

This mission report was prepared within the EU Twinning Project "Further Support to the Reform of the Statistics System in Bosnia and Herzegovina". It was the fifth mission to be devoted to Agriculture Statistics, Component 2 of the project.

The purposes of the mission were:

- Review of the sample, details of its design and some information on data collection
- Data editing on the basic level, focusing especially on the detection of the (significantly) outlying values
- Response analysis, calculation of the response rates and coverage rates on the entity level
- Calculation of non-response adjustment weights
- Calculation of final weights
- Estimation and analysis of key variables
- Deciding on the practical procedure for the estimation of the under coverage ratio, the over coverage ratio, the population size, and their relative standard errors

The consultant would like to express his/her thanks to all officials and individuals from Bosnia and Herzegovina met for the kind support and valuable information, 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, BHAS, FIS, RSIS, CBBH, Statistics Denmark, Statistics Finland, Statistics Sweden and The Italian National Institute of Statistics.

2 Assessment and results

2.1 Review of the sample

Sampling design:
- Stratified cluster sampling.
- Strata: Entities (3) and type of settlement (2, urban/other).
- Enumeration areas (hereinafter EA) were used as PSUs (*Primary Sampling Units*)
- PSUs inside the strata were selected by the PPS (*Probability proportional to size*) method, using sum of two standardised variables (UAA (*Utilized agriculture area*) and LSU (*Livestock units*)) as the size variables.
- All units inside the selected cluster were to be surveyed.

Data collection in the field:
- The samples in rural strata were divided in 2 sub-samples.
  - o The first sub-sample was used to implement the „door-to door" approach using maps as well as list of buildings (no matter if they were occupied or not), created on the basis of the Census 2013 data. In this case, the all new agriculture and nonagricultural holdings, found in the field were also

4

surveyed. For this sample, they used the short form of questionnaire, including status questions and few questions on agriculture activity. In the phase of setting up the survey database, a bit different approach was used by Entities. In Entities 1 and 3 (Federacija and District Brčko) the database survey lists also included all unocupied buildings that were in the initial lists, while in Entity 2 (Republika Srpska) in the CAPI (*Computer-Assisted Personal Interviewing*) application and the final database, were entered only data for buildings (and households) in which the households (agricultural or nonagricultural households) exists or existed according to the results of PC 2013. Data for all new households found within the EA are entered into the application and final database.

- o The second sub-sample used only „list approach", where the term list refers to list of households in rural EA and list of agricultural households in urban EA. For urban EA only households engaged in agriculuture in Census in BiH in 2013 were on this list (not all as in sub-sample 1). This list was updated with data from Ministry of agriculture from 2022 both in Institute for Statistics of FBiH (FIS) and RSIS. In the case of Entities 1 and 3, the lists only contained the households, i.e. only occupied buildings. In the case of Entity 2 or Republika Srpska, in the frame of second sub-samle used two diferrent approach. In the rural EA, RSIS used „List of husehulds", that contained all households from results of PC 2013 without updates from administrative sources and the ability to identify new households in the EA. In the urban EA, RSIS used „List of agricultural husehulds", that contained all households from results of PC 2013 with agricultural activities updated with farms from administrative sources (MoA). The units that by the data from the short questionnaire exceeded the threshold, were also surveyed with the long questionnaire. Here we have to meantion that the data of long questionnaire was not treaed at all during this mission.
- o In Entitiy 1, 19 out 112 enumeration area were not surveyed at all. The reason was the lack of interviewers that were on disposal in the data collection period. We treated these cases as the non-reponse on the „EA level". As will be explained later, additiona non-response correction factor was calculated to overcome this problem.

2.2 Data integration

The first two days of the mission were mostly devoted to the problems with the data that came from two different data collection systems. The main reason for the problems with the data was the fact that the data from two different systems were not merged on time. Further on, there was a significant error in the submitted table that the entities were supposed to fill out. Harmonisation of the variable labels and alignment of the response status categories, were just a part of the process of setting-up the correct data file. In parallel with the these

5

"integration activities", the rough editing of the raw data was carried out. Few extreme outliers were corrected as well as few obviously wrong response statuses.

2.3 Response analysis

On the basis of the collected response statuses, all sample units were classified into three categories: Responses (1), Non-responses (2), Out of scope units (3). Additional categorisation was then implemented, by using the question *Does your household carried out any agriculture activity*. All the units that responded No to this question, were by this second categorisation treated as "out of scope units", since they are inadequate from the agriculture point of view. In the following tables we present the shares by both categorisations.

Table 1: Shares of response statuses by categorisation 1

Domain	Domain value	Status code	Rate
Total		1	71,5%
Total		2	9,7%
Total		3	18,8%
Entity	1	1	75,5%
Entity	1	2	9,3%
Entity	1	3	15,1%
Entity	2	1	66,6%
Entity	2	2	9,7%
Entity	2	3	23,7%
Entity	3	1	64,5%
Entity	3	2	13,3%
Entity	3	3	22,2%
Settlement Type	Urban	1	70,1%
Settlement Type	Urban	2	11,4%
Settlement Type	Urban	3	18,5%
Settlement Type	Rural	1	71,5%
Settlement Type	Rural	2	9,6%
Settlement Type	Rural	3	18,9%

Table 2: Shares of response statuses by categorisation 2

Domain	Domain value	Status code	Rate
Total		1	55,0%
Total		2	9,7%
Total		3	35,3%
Entity	1	1	61,7%
Entity	1	2	9,3%
Entity	1	3	29,0%

6

Entity		2	1	48,4%
Entity		2	2	9,7%
Entity		2	3	41,9%
Entity		3	1	31,9%
Entity		3	2	13,3%
Entity		3	3	54,8%
Settlement Type	Urban	1		41,7%
Settlement Type	Urban	2		11,4%
Settlement Type	Urban	3		46,9%
Settlement Type	Rural	1		55,6%
Settlement Type	Rural	2		9,6%
Settlement Type	Rural	3		34,8%

2.4 Calculation of weights

The weights were calculated through the following steps:

- **Design weight.** The weights that were constant inside each of the PSUs, were determined as the reverse value of the selection probability. Selection probabilities were taken from the sampling phase.
- **Non-response correction.**
  - Factor for the non-response at the "EA level". This factor was in fact applicable only in Entity 1, where 19 EAs were not surveyed at all. This correction factor was calculated separately for each stratum h by using the following formula:

    $$NR_{CL}^h = \frac{\sum_{i=1}^{C} w_{di}^h}{\sum_{i=1}^{R} w_{di}^h}, \text{ where}$$

    C...number of all clusters inside stratum h
    R...number of inside stratum h that were surveyed
    w_{di}^h...design weight of cluster i inside stratum h

  - Non-response correction factor for the non-response inside the surveyed clusters. We calculated two different versions of this factor. The first one was calculated in order to be used for the units that were selected from the list, while the second was intended to be used for the "new units", i.e. the units that were not on the list, but were found in the field by using the "door to door" principle. The factors were calculated inside each stratum h by using the following formulas:

    $$NR_h^1 = \frac{\sum_{i=1}^{n_h} x_i}{\sum_{i=1}^{r_h} x_i + \sum_{i=1}^{p_h} x_i}, \text{ where}$$

    n_h...number of sample units in stratum h
    r_h...number of responses in stratum h
    p_h...number of "out of scope" units in stratum h
    x_i ... value of the auxiliary variable (variable that is available for each unit in the sample) for unit i. We took UAA from the Census 2013 as the auxiliary variable.

$$NR_h^2 = \frac{\sum_{i=1}^{n_h^1} x_i + \sum_{i=1}^{s_h} x_i}{\sum_{i=1}^{r_h} x_i + \sum_{i=1}^{p_h} x_i + \sum_{i=1}^{s_h} x_i} \cdot k_{s1}, \text{ where}$$

n_h^1...number of sample units in stratum h, taking into account only sub-sample 1

r_h...number of responses in stratum h, taking into account only subsample 1

p_h...number of "out of scope" units in stratum h, taking into account only subsample 1

s_h...number of "new units" in stratum h

x_i ... for the new units, the UAA collected in the Pilot Agricultural Census

k_{s1}... design correction factor for sample 1, calculated as:

$$k_{s1} = \frac{\sum_{i=1}^{n_h} w_{di}^h}{\sum_{i=1}^{n_h^1} w_{di}^h}, \text{ where}$$

The final weight was calculated as

- $w_i^h = w_{di}^h \cdot NR_{CL}^h \cdot NR_h^1$ for the "list units"
- $w_i^h = w_{di}^h \cdot NR_{CL}^h \cdot NR_h^2$ for subsample 1(including "new units")

2.5 Results of the analysis

Due to the limitation of the available time, only some basic analysis were carried out. We were focusing mainly on estimation of the under coverage and over coverage in the census lists.

- Over coverage

  In the frame of "over coverage estimation", we aim at estimating the share of units that were according to the "Census 2013 list" eligible to be surveyed, but are according to the current survey data not eligible anymore (are out of scope).

  In the first table, we present the response categories for the units that according to the Census 2013 carried out (at that time) some agriculture activity. The "NO category" includes the units that were ineligible from the data collection point of view (e.g. building empty) as well as the units that responded that they don't have any agriculture activity.

Table 3: Shares of maintenance of agricultural activity

Entity	Still have agriculture activity	
	YES	NO
1	72%	28%
2	65%	35%
3	60%	40%

For the second table, we took two key variables from the "Census list" (UAA and Cattle) and estimated what is the share of the total from 2013 that is now "lost" (due to the out of scope units in the survey).

Table 4: Shares of maintenance of agricultural activity by key variables

Entity	UAA	Cattle
1	33,8%	21,0%
2	33,7%	31,4%
3	48,4%	30,4%

- Under coverage
  In order to estimate the under coverage of the Census lists, we used the "new units", which were surveyed by using the "door to door" approach in subsample 1. Here we want to point out that "new" doesn't necessary mean that the holding didn't exist in 2013, but could just from different reasons be dropped out of the list. In the future, some further analysis should be carried out on these new units, to find out what is the reason for this relatively high number of such units.
  In the below table we present the share of the new units in the estimated population totals. We here point out that the totals of new units were estimated only based on subsample 1, while the totals for "old" units were estimated by using the total sample.

Table 5: Shares of New units, estimated by subsample 1

Entity	Number of units	UAA	Cattle
1	23,2%	23,4%	14,9%
2	4,5%	9,9%	6,2%
3	2,1%	0,5%	0,0%

3. Conclusions and recommendations

The key conclusions:
- A lot of work was spent for the integration of data from two different systems for data collection (IST and CSPro). In addition, it turned out that the questionnaires that were used by the three institutions were not fully harmonised.
- The data that we worked with were more or less raw. We were only capable of carrying out some basic editing, focusing on outlying values and on ensuring consistency between collected data and response statuses.
- Two different set of weights were calculated. The first set is to be used for the extrapolation of the whole sample, while the second set is to be used only for the extrapolation of the first subsample data.
- So far, the weights were used in order to estimate the under- and over coverage of the Census 2013 lists.

Recommendations:
- For the Census, the full harmonisation of the questionnaire that will be used in different entities should be ensured in the preparation phase.
- The usage of only one system for questionnaire creation and data collection should be considered. With such unique system, the problems with data integration should be significantly reduced.
- Some additional work should be put into the data editing, especially into data editing of the long questionnaire, the data of which were during this mission not tackled at all.
- Additional analysis of the so-called "new units" should be carried out. The focus should be on the reasons, why these units were not included in the Census lists.
- The currents results (high over- and under- coverage rates) indicate that the door to door approach in rural areas should be the preferable one, however some further analysis should be carried out before the final decision.

4. What to do before the next mission for the BC Counterpart

Action	Deadline	Responsible person
Some further analysis on "new" units		
The proper integration of the "long questionnaire" data from two different data collection systems		

Annex 1. Terms of Reference for the current mission

Terms of Reference

EU Twinning Project BA 17 IPA ST 01 20

Component 2 - Agriculture Statistics
Period: 19 – 22 December 2022
(23. December: Final input and reporting)

Venue: Agency for Statistics of BiH, Zelenih beretki 26
71 000, Sarajevo
Bosnia and Herzegovina

Activity 2E1: Analysis of the Pilot Agricultural Census Results

1. Mandatory result

The component on Agricultural Statistics has three main objectives:

- 2.1 Master Plan for Agriculture statistics for the period 2020-2030 developed and adopted (with clearly defined objectives, timetable of implementation for the implementation of the objectives and precisely defined responsibilities of all stakeholders.)
- 2.2. Cooperation among all stakeholders improved
- 2.3. Protocol(s) on cooperation and exchange of data among all stakeholders adopted

Indicator / Relevant Milestones / Internal deadlines:

- Master Plan for Agriculture statistics in BiH available on web
- Identification of stakeholders at national and entity level
- Responsibilities of each stakeholders clarified
- Overview of currently available data both administrative and from surveys
- Development of roadmap for implementation
- Development of initial budget for agricultural census
- Definition of SRPG ~Statistical Register of Agricultural Holdings
- Development of high-level cooperation agreements / Memorandums of Understanding
- Protocols describing specific exchange of data

11

Purpose of the activity

- Review of the sample, details of its design and some information on data collection
- Data editing on the basic level, focusing especially on the detection of the (significantly) outlying values
- Response analysis, calculation of the response rates and coverage rates on the entity level
- Calculation of non-response adjustment weights
- Calculation of final weights
- Estimation and analysis of key variables
- Deciding on the practical procedure for the estimation of the undercover age ratio, the over coverage ratio, the population size, and their relative standard errors

2. **Expected output of the activity**

- Sample design and data collection strategy reviewed
- The most significant outliers from the non-exhaustive parts detected and corrected
- Coverage and response quality indicators calculated
- Non-response adjustment weights calculated
- Final weights calculated
- Design of estimation strategy
- Grossing up procedure of the key variables and analysis of estimations defined
- Coverage indicators estimated

Annex 2. Persons met

BHAS:
Fahrudin Subotić
Ljubiša Eskić
Biljana Nogo
Ljubica Marić
Andrea Mehmedović
Lidija Lukić
Jasna Isaković
Neno Svjetlanović

RSIS:
Vedrana Dejanović
Lazo Šegrt

FIS:
Elmina Ramić-Lukač
Ismir Melez
Asmira Selimovic
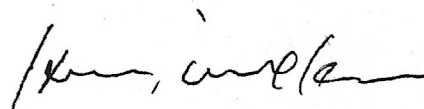Amel Sikirić

RTA Team:
Niels Madsen
Larisa Muslimovic

Signatures

For the approval of the contents of this report, representatives from BHAS, FIS and RSIS as well as MS experts and the RTA sign here:

Date:

Component leader, BHAS

Component leader, RSIS

MS Expert

Component leader, FIS

RTA

13