TWINNING CONTRACT

JO/13/ENP/ST/23

Strengthening the capabilities of the Department of Statistics in Jordan

MISSION REPORT

on

Activity: 2.6 Test run of the new sampling techniques

Mission carried out by
Mr Nigel Stuttard, NI-CO
Mr Mārtiņš Liberts, Central Statistical Bureau of Latvia

18-22 January 2015

Version: Final

Expert contact information

Nigel Stuttard
NICO
Landmark House
5 Cromac Quay
Belfast, BT7 2JD
United Kingdom
Tel: +44 1708 221318
Email: nanddstuttard@hotmail.co.uk

Dr Mārtiņš Liberts
Central Statistical Bureau of Latvia
Lāčplēša Street 1
Rīga, LV-1301
Latvia
Tel: +371 67 366 833
Email: martins.liberts@csb.gov.lv

Table of contents

1. General comments 4
2. Assessment and results 4
3. Conclusions and recommendations 9
Annex 2. Persons met 15

List of Abbreviations

DoS	Department of Statistics of Jordan
ToR	Terms of Reference

1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capabilities of the Department of Statistics in Jordan".

The purposes of the mission were:

- Based on the recommendations provided in the previous activities, practical application of sampling methodologies on different types of surveys (agricultural, economic, household). With a special focus on:
    - Sampling design (Redesign of the Agriculture Capital Formation Survey)
    - Sampling errors (Estimation for various agricultural, economic, households surveys)

- If the time allows it, the following topics will be discussed as well.

    - Imputation
    - Weighting
    - Non-response
    - Seasonal adjustment
    - Discussion on software solutions

2. Expected output of the activity

- Recommendations prepared on the redesign of the agriculture capital formation survey.
- Recommendations prepared on estimation of sampling errors in the various surveys.
- The implementation of recommendations regarding sampling methodologies, prepared after previous activities in the component, are discussed
- Recommendations prepared on how to solve the remaining challenges
- Transfer of the Latvian and British and in general the European Union, experience in sampling methodologies

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 Jordan and which highly facilitated the work of the consultant.

The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, DoS, the Central Statistical Bureau of Latvia or NI-CO.

2. Assessment and results

The mission was carried out mainly through discussions between the experts and the DoS staff. The experts used these discussions as a base for interrogating the DoS survey data and developing the software solutions.

Capital Formation Survey

The focus of discussion with the Agriculture Section was The Capital Formation Survey (CFS). This survey covers arable farming in both the Jordan valley and the uplands, together with livestock farming in both the organised and unorganised sectors. It aims to collect data about the capital stock - buildings, animals, plantations, wells, reservoirs, machinery and equipment - at the beginning of the year and the change in stock over the year. The data is collected in a single visit to each farm with the survey conducted in the different areas at different times.

The main use of the survey is to derive estimates for the change of stock, i.e. capital formation over the year, for use in the national accounts. Estimates for the level of stock are initially based on census data with estimates for the years following the census derived by applying the changes to the estimates for the base year. The data is augmented with data from other sources when that data are seen as more reliable.

Separate estimates for the capital investment in Buildings, Wells and Reservoirs, and Machinery and Equipment are derived in a similar ways by using data from the survey to measure the change through development, purchases, sales and loss adjusting the stock for depreciation. The measurement of change is based on very few observations, for example only 14 out of 3,975 units reported a change in buildings, with none reporting either sales or purchases of new buildings. This means that estimates are likely to be highly variable between years, especially as the weights for different farms vary considerably, from 1 to 56. In addition, as can be seen from the table below, depreciation has a large impact on final estimate of capital investment, for example depreciation is 8 times as large as investment in buildings. This may be a true picture of what is happening in the economy but if depreciation is making such a large impact on capital investment it is important that the estimates of the rates of depreciation are as accurate as possible. We were unable to find out how these rates of depreciation were estimated during the mission but we would recommend that work is carried out to decide whether these estimates are the best available.

	Stock (end 2012)	Investment in 2013	Stock (end 2013)	Depreciation
Buildings	372,665,300	2,456,165	375,121,465	11,253,644
Wells and reservoirs	116,262,887	5,386,492	121,649,379	12,164,938
Buildings and equipment	276,878,765	3,478,229	280,356,994	28,303,092

Estimates for capital formation in livestock were slightly more complicated but logical. We had some difficulty arriving at the same estimates as DoS, especially for cattle. This may well have been because the weights we were using but it did give rise to concern about both version control - when corrections/adjustments are made to a spreadsheet, do they necessarily feed through to other versions of the same spreadsheet? - and documentation. A further concern is that the valuation of livestock is provided by the respondent. While one would expect the farmer to know what his stock is worth, it would be interesting to know the extent to which their valuation varied and whether such variation is justified.

Estimates for capital formation for fruit trees is calculated by multiplying the number of new trees, estimated from a more comprehensive survey, by the cost per tree, estimated from the CFS, and adding in additional investment in new trees. In 2013 only 67 of the 1,415 survey units reported buying new trees and these responses were used to estimate the cost of 12 different types of tree. Even more worrying, only two respondents reported additional investment in new trees. Again this means that estimates are based on few observations and are therefore likely to be of low quality and very variable.

Crops Production Survey

The aim was to develop a procedure for the estimation of sampling errors in Crops Production Survey (CPS). The survey data from 2013 for two seasons were used as a test data for the development of the procedure.

Two stage sampling design is used for the CPS. Villages are the primary sampling units and farms are the second stage sampling units. Villages are sampled using stratified simple random sampling where stratification is done by governorates and strata variable.

Sampling errors were estimated for the estimate of the total production for the field crops, open field vegetables, covered vegetables and fruit trees (for each season separately). The estimates of sampling errors were done for the kingdom and for each governorate.

Sampling errors were estimated using the Ultimate Cluster Method. The sampling error was expressed as the coefficient of variation. The code for sampling error estimation was written using IBM SPSS Statistics. The alternative code was written in R using the functions from the package "vardpoor". The alternative code was made to test the correctness of the results.

The survey data and weights were analysed during the estimation of the sampling errors. We found that the calculation of weights for the primary sampling units does not match with sampling design used for sampling the primary sampling units. The primary sampling units are sampled using stratified simple random sampling, however the weights for sampled primary sampling units are computed according to the stratified πps sampling designs (sampling design with sampling probabilities proportional to the unit size).

We talked with DoS employees about how to produce the weights for the sampling design. It was agreed to recalculate the weights and to run the code for sampling errors with correct weights. The estimates of the coefficients of variation derived during the mission should be considered only as an example. They cannot be used for precision evaluation until the reweighting has been done and the estimation of the coefficients of variation has been repeated with correct weights.

Agricultural Prices Survey

The aim was to develop a procedure for the estimation of sampling errors in Agricultural Prices Survey. Partial survey data from Jordan Valley were used as a test data for the development of the procedure.

Single stage stratified sampling of farms is used in Jordan Valley. Sample allocation is computed as proportional to the population size in each stratum. Two stage sampling design is used in the uplands.

The parameters of interest are farm gate price and market price for different vegetables and fruits. The type of statistic is ratio of two totals. Taylor linearisation was applied to derive approximate variance for the estimates of ratio.

Sampling errors were estimated using the classical variance estimator for the stratified simple random sampling design. The sampling error was expressed as the coefficient of variation. The code for sampling error estimation was written using IBM SPSS Statistics. The alternative code was written in R using the functions from the package "vardpoor". The alternative code was made to test the correctness of the results.

The survey data and weights were analysed during the estimation of the sampling errors. We concluded that the sampling weights should be calculated according to the sampling design and weights should be used for the price estimation. It was not possible to estimate sampling errors for uplands as sampling weights were not available.

We discussed with DoS employees how to calculate sampling weights. They agreed to calculate the weights for the survey according to the sampling design. The estimates of the coefficients of variation derived during the mission should be considered only as an example because only partial survey data from the Jordan Valley were available during the mission. Complete survey data for the Jordan Valley should be prepared. The same code for sampling errors can be used on the complete dataset to derive the precision estimates. The code should be amended for the estimates of uplands as different sampling design is used there.

Economic Survey on Transport

The aim was to develop a procedure for the estimation of sampling errors in Transport Survey (TS) for the unorganised sector. Full enumeration is used for the organised sector, so sampling errors are not observable for the organised sector. The survey data from the second quarter of 2014 were used as a test data for the development of the procedure.

Stratified simple random sampling is used for the TS. Stratification variables are governorate and type of vehicle (taxi and services, micro buses, buses, pickups, trailers and refrigerators, and tankers). Transport vehicles are the sampling units. The survey is done each quarter. The sampling fraction is 1.5%. The reference period is one month of the quarter. Monthly survey data are extrapolated to quarterly data by multiplying the monthly figures by three.

Sampling errors were estimated for the estimate of the total gross output and intermediate consumption. The estimates of sampling errors were done for the total unorganised sector, separately for passenger and freight vehicles and by type of vehicle (six types).

Sampling errors were estimated using the classical variance estimator for stratified simple random sampling. The sampling error was expressed as standard error, the coefficient of variation, and confidence interval (at 95% confidence level). The code for sampling error estimation was written using IBM SPSS Statistics. The alternative code was written in R using the functions from the package "vardpoor". The alternative code was used to test the correctness of the results.

The resulting estimates of sampling errors are shown in the following tables. The prepared code can be used for other periods of the survey. It is advisable to derive the sampling error estimates for several quarters to evaluate the precision of the survey.

Standart errors and coeficients of variation

period: 2 Q 2014

		variable					
		Gross output			Intermediate Consumption		
		y_total	se	cv	y_total	se	cv
Type of vehicles	Passenger Vehicles	163,843,102	1,721,122	1,1	74,416,069	2,169,405	2,9
	Taxi and Services	102,807,594	1,513,060	1,5	53,206,629	2,131,502	4,0
	Micro Buses	39,068,735	739,828	1,9	14,317,687	377,160	2,6

	Buses	21,966,773	354,349	1,6	6,891,753	144,109	2,1
	Freight Vehicles	201,626,506	4,480,522	2,2	75,092,763	1,446,946	1,9
	Pickups	29,926,576	1,024,441	3,4	11,678,473	298,310	2,6
	Trailers and Refrigerators	153,885,992	4,277,194	2,8	55,776,781	1,349,424	2,4
	Tankers	17,813,937	855,104	4,8	7,637,509	428,622	5,6
	Total	365,469,608	4,799,722	1,3	149,508,833	2,607,675	1,7

Confidence intervals (at 0.95 confidence level)

period: 2 Q 2014

		variable					
		Gross output			Intermediate Consumption		
		y_total	ci_lo	ci_up	y_total	ci_lo	ci_up
Type of vehicles	Passenger Vehicles	163,843,102	160,469,704	167,216,501	74,416,069	70,164,035	78,668,104
	Taxi and Services	102,807,594	99,841,996	105,773,192	53,206,629	49,028,884	57,384,374
	Micro Buses	39,068,735	37,618,672	40,518,798	14,317,687	13,578,454	15,056,921
	Buses	21,966,773	21,272,249	22,661,297	6,891,753	6,609,300	7,174,207
	Freight Vehicles	201,626,506	192,844,684	210,408,328	75,092,763	72,256,750	77,928,777
	Pickups	29,926,576	27,918,672	31,934,481	11,678,473	11,093,785	12,263,162
	Trailers and Refrigerators	153,885,992	145,502,692	162,269,293	55,776,781	53,131,910	58,421,653
	Tankers	17,813,937	16,137,934	19,489,941	7,637,509	6,797,409	8,477,608
	Total	365,469,608	356,062,153	374,877,064	149,508,833	144,397,789	154,619,877

Confidence interval covers the true population value with probability 0.95

3. Conclusions and recommendations

Capital Formation Survey

It is clear that estimates of capital formation in these different facets of agriculture are based on very few observations in the field and are therefore likely to vary considerably between years. Investment is always difficult to pick up as establishments do not necessarily make major investments every year and this may well be even more the case for well established farms. One way of overcoming this problem would be to combine several years of data. As well as providing a more robust annual estimate, it would reduce the variation between years as there would be a considerable overlap between the samples used in the annual estimates. Rolling annual estimates should be based on at least 5 years of data.

However, it seems likely that the survey is not well designed to pick up investment. Respondents are asked to report on their holdings at the beginning of the year and on any changes that have taken place over the year. Any survey which relies on respondents recalling information is likely to problematic. It would be better to ask about the situation now and then return 12 months later and ask again. This suggests that the best design for a survey would be to have a rolling panel with, say, each farm staying in the survey for 5 years and 20% of the sample being renewed each year. During the annual visits, respondents would be asked about the situation now and the changes over the last twelve months. We understand that it may be possible to introduce computer aided interviewing and this would improve the accuracy of such a survey considerably as the interviewer and responder could explore together what had happened over the last year.

Even though it was the purpose of this activity to redesign the CFS, it wasn't possible to meet this part of the ToR. However, recommendations based on the experts best experiences were provided to DoS. See above.

Crops Production Survey

The sampling weights for the primary sampling units should be calculated according to the sampling design used for sampling (stratified simple random sampling in this case). The estimates of population parameters with appropriate weights should be derived and compared with the current estimates. The code for sampling error estimation has to be used with the appropriate weights to derive appropriate sampling error estimates.

Agricultural Prices Survey

The weights should be calculated for the sample in uplands. The full survey data has to be prepared for the precision estimation. The code developed during the mission can be used for the Jordan Valley (where stratified simple random sampling is used). The calculation of sampling errors for uplands has to be amended according to the two stage sampling design (see the code for the Crops Production Survey).

Economic Survey on Transport

The code for sampling error estimation has to be used to derive precision estimates for other quarters. The precision estimates from several quarters (four or eight last quarters available) can be used to evaluate the precision of survey and this information can be used to reallocate the sample size by strata if necessary.

Action	Deadline	Responsible person
In the short term, estimates of capital formation in agriculture to be based on combining data of the latest five years using the current methodology.	In time for the calculation of national accounts for 2014.	Directorate of Agricultural Statistics
Steps are taken to improve the documentation and version control for agriculture statistics.	With immediate effect.	Directorate of Agricultural Statistics
DoS develop a rolling panel survey designed to collect capital formation in agriculture.	When DoS work plan allows	Directorate of Agricultural Statistics
Recalculation of sampling weights for the Crops Production Survey.	When DoS work plan allows	Directorate of Agricultural Statistics
Calculation of sampling errors for the Crops Production Survey using the code delivered with the correct weights.	When DoS work plan allows	Directorate of Agricultural Statistics
Calculation of sampling weights for the Agricultural Prices Survey.	When DoS work plan allows	Directorate of Agricultural Statistics
Preparing a complete survey data set for Agricultural Prices Survey in Jordan Valley. Calculation of sampling errors for the Agricultural Prices Survey in Jordan Valley using the code delivered.	When DoS work plan allows	Directorate of Agricultural Statistics
Calculation of sampling errors for Economic Survey on Transport (for several quarters) using the code delivered.	When DoS work plan allows	Directorate of Economic Surveys
Evaluation of the precision of Economic Survey on Transport using the sampling error estimates from several quarters.	When DoS work plan allows	Directorate of Economic Surveys

Terms of Reference

EU Twinning Project JO/13/ENP/ST/23

18-22 January 2015

Component 2: Sampling techniques

Activity 2.6: Test run of the new sampling techniques

0. Mandatory results and benchmarks for the component

- Improve the capacity of DoS staff to understand and apply modern sampling techniques (Apr 2015)
- Assessment report on current situation (Jan 2013)
- Provide inputs to the design of surveys (Aug 2014)
- Conduct a training course in seasonal adjustment (Oct 2014)
- Give recommendations on how to deal with weights, imputation, non-response and sampling errors (Apr 2015)

1. Purpose of the activity

- Based on the recommendations provided in the previous activities, practical application of sampling methodologies on different types of surveys (agricultural, economic, household). With a special focus on:
  - Sampling design (Redesign of the Agriculture Capital Formation Survey)
  - Sampling errors (Estimation for various agricultural, economic, households surveys)
- If the time allows it, the following topics will be discussed as well.
  - Imputation
  - Weighting
  - Non-response
  - Seasonal adjustment
  - Discussion on software solutions

2. Expected output of the activity

- Recommendations prepared on the redesign of the agriculture capital formation sourvey.
- Recommendations prepared on estimation of sampling errors in the various surveys.
- The implementation of recommendations regarding sampling methodologies, prepared after previous activities in the component, are discussed
- Recommendations prepared on how to solve the remaining challenges
- Transfer of the Latvian and British and in general the European Union, experience in sampling methodologies

3. Participants

DoS
Mrs Ghaida Khasawneh, Head of Sampling Division *(Component Leader*)

Staff from the Directorate of statistics studies
Staff from the Directorate of Economic surveys
Staff from the Directorate of Agricultural statistics

MS experts
Mr Martins Liberts, Deputy Head of Mathematical Support Division, Central Statistical Bureau of Latvia
Mr Nigel Stuttard, NI-CO

Programme for the mission

Time		Place	Event	Purpose / detail
Sunday, morning	08.30 – 10.00	Hotel /DoS	Meeting with RTA	To discuss the programme of the week
Sunday, morning	10.00 – 12.00	DoS	Meeting with BC Component Leader and BC Experts	Discussions of the week's programme Meeting with staff from the Directorate of Agricultural Statistics and the Directorate of National Accounts regarding the redesign of the Agriculture Capital Formation Survey
	12.00 – 01.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Sunday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Continued.
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Monday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 12.00		Meeting with BC Component Leader and BC Experts	Continued.
	12.00 – 01.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Monday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Continued.
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Tuesday, morning	08.30 – 10.00	DoS	Meeting with BC Component Leader and BC Experts	Meeting with the staff from the Directorate of Economic Surveys regarding estimation of sampling errors.
	10.00 – 12.00			Break / Preparations / Report writing
	12.00 – 01.00		Break / Preparations / Report writing	

Tuesday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Continued.
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Wednesday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 12.00		Meeting with BC Component Leader and BC Experts	Continued.
	12.00 – 01.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Wednesday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Continued.
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Thursday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 12.00		Meeting with BC Component Leader and BC Experts	Meeting with the staff from the Directorate of Statistical Studies regarding estimation of sampling errors in the household surveys.
	12.00 – 01.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Thursday, afternoon	01.00 – 02.30	DoS	Meeting with BC Component Leader and BC Experts	Continued.
			Ad-hoc meetings	Final clarifications with BC Experts, preparation of report and presentation for BC Project Leader
Thursday, afternoon	02.30 – 03.00	DoS	Meeting with BC Component Leader	Presentation for BC Project Leader
Thursday, afternoon	03.00 – 04.00	DoS	Debriefing with BC Project Leader	Conclusions and decisions and their consequences for the implied work programme for BC Experts

Annex 2. Persons met

DoS:

Mr Abed Wadood Matouk; Project Leader

Methodologies and Statistical Techniques Directorate:
Mrs Ghaida Khasawneh, Head of Sampling Division (component leader)
Mr Abdalnaser Jariri, Statistician of Sampling Division

Economic Directorate:
Mr Mohammad Fathi

Agricultural Directorate:
Mr Mohammad Tawalbih
Mr Mohammad Yahya Al Jawarneh, Statistician of Crops Production Division
Mr Abdelnaser Mohamad Obeidat

Directorate of National Accounts:

RTA Team:

Mr Thomas Olsen, Resident Twinning Adviser (RTA)
Ms Christine Salman, RTA assistant
Mr. Mohammad Junaidi, Interpreter