The household budget and expenditure data collection module (IOF 2014/2015) within a continuous multipurpose survey system (INCAF)

Report from a forth short term mission to the National Statistical Institute of Mozambique, Maputo Mozambique

9-20 June 2014

within the frame work of the

AGREEMENT ON CONSULTING ON INSTITUTIONAL CAPACITY BUILDING, ECONOMIC STATISTICS AND RELATED AREAS

between
INE and Scanstat

Lars Lundgren

INSTITUTO NACIONAL DE ESTATÍSTICA

Ref: Contract DARH/2008 /004

Lars Lundgren
larsalundgren@hotmail.com

Contents

Summary

INE has decided to modernise its household survey system by introducing a continuous multi-purpose survey (INCAF) started the INCAF 1 July 2012, but due to funding problems only the first quarter was fully processed. All resources have now been focused on the household budget module IOF (Inquérito ao Orcamento Familiar), which can be seen as a rotating module. The experience from INCAF has been used in the planning of IOF and a pilot was conducted in April/May 2014.

The INCAF was evaluated by an independent Italian team and the pilot was evaluated by the INCAF/IOF team. Found problems have been documented and attended.

Problems and questions	Findings and recommendations
Strategic considerations at questionnaire design	International comparability and comparability backwards and forward. Following NA and CPI rules and previous welfare/ poverty standards.
Simplicity	The more complex, the less quality. The IOF 2014/15 will collect more data than previous IOF. Labour force and environment are expanded.
No quantities	Quantities will still be collected, though most errors in the pilot were linked to this area and the data are of very limited use, if any.
4x4 methodology	Each household will be measured all quarters and all types of weeks of the month, to avoid misclassification due to seasonal variation and monthly variation related to salary payments (contrary to the pilot report).
Full year	In the pilot semi-durables were measured for one month each quarter. This will be changed to 3 months. At the first visit the interviewer will ask for past 3 month and at the following quarterly visits about expenditures since last visit. Besides reducing the standard errors it will make the start of the periods more clear.
Partial non-response	Moved households will be imputed by replacing households, but marked for flexible handling.
Panel rotation	No panel rotation will be used during the IOF-module.
Sub-samples	Sub-samples should be used when smaller samples are sufficient, e.g. if monitoring subjective well-being.
Lump-sum	The lump-sum methodology has not been included in IOF, but should be tested on small sub-sample (500 households) during the second half of the survey.
Short vs. long lists of items	IOF is aiming to collect all consumption. It is done by asking people of what they have bought since the last visit (supplemented by diary keeping), not asking item by item from a list. The size of the list should therefore not have an impact on the results. IOF 2014/15 will have longer lists than previous IOF.
COICOP	The codes are much extended for clothing. For food outside home distinction is not made by type of venue (restaurant, café/bar or canteen) but type of consumption (food, non-alcoholic beverages or

	alcohol). A separate questionnaire could be used for a small sub-sample to estimate more details.
	The pilot had too many 999 999 as COICOP code. Recoding at INE is costly and time consuming and risks the quality as central staff may lack enough information. The item lists should be updated to cover most of these items, even in local languages. The training and supervision should also address this problem.
Individual questionnaires	Individual questionnaires will be used to better collect individual expenditure, particularly food outside home and transport expenditures.
Out-layers	The pilot had a second-hand car for 30 000 000 Mt. CAPI should give a signal for potential outliers.
Data transactions	Data are transferred by Bluetooth from the interviewers to the supervisors and from the supervisors to both regional offices (as back-up) and to INE HQ as RAR-files. Handshaking and encryption is not used.

Unsafe areas (Sofala) and coming election	Some areas may be too unsafe to visit. If it is more permanent local interviewers can be used, but otherwise the areas will be excluded. The documentation will declare that the survey is representative except for those areas.
	Elections have been going on during previous household survey without implications. The interviewers are trained to be neutral and will already have visited all households before the election. Accidents outside control of the survey can always happen.
Poverty lines	Regional poverty line is much recommended as the food culture differs in such a tall country. People's competence to get most energy for the money is more relevant imposing the same food basket on everyone. Household poverty lines would even be better.
Multidimensional welfare/poverty	Like previous IOF, the new survey will generate statistics in many areas relevant for welfare: employment, education, health, housing, services, possessions, etc. It is up to the analysts to decide if those data will be combined into a composite index. The main problem is the lack of weights.
	A better alternative is to use a subjective well-being approach focusing on service performance and quality of life (including good governance, human rights and participation).
Tabulation and dissemination	The first quarter will be tabulated similarly as the first quarter of INCAF, i.e. focusing on labour force statistics.
	It is also important to use the first quarter to evaluate the quality and to improve processes, as well as planning the future reports in more details.
	The full year will be tabulated similarly as the previous IOF, focusing on socio-economic and consumption statistics. Specific module, as Tourism will be tabulated by relevant departments and MPD will be responsible for poverty analysis.
	Some improvements are proposed.
	The first quarter is expected to be disseminated in November 2014 and the full year in November/December 2015.
Risks	With many visits to the households, the data collection is not expected to be at risk. As in previous IOF the main risks will occur in data processing. A robust database is needed.

The consultancy

The aim of the consultancy was to optimize the implementation of INCAF and IOF, based on the experiences from the pilot. The proposed activities were:

1. Assist INE by giving recommendations for the new INCAF with income and expenditure module (IOF);
2. Support INE in the tabulation plan and its reports;
3. Share experience on conducting continuous multi-purpose surveys, including preparation and presentation of the results;
4. Assist in tuning the methodology by discussing with stakeholders regarding the use of the results;

The consultancy was implemented during the training of the interviewers in Boane and the consultant also stayed in Boane most of the time to support in upcoming problems. Meetings with the donors and the PAM-group were held in Maputo during two separate days. The users are only focusing on the IOF to get results for the Millennium Development Goals and the strategies for poverty reduction. The consultant also participated in three meetings for the Modernisation project, as INCAF will be the main object for the project. This draft report was discussed at the end of the mission. The full ToR are found in annex 1.

The pilot

The pilot tested the whole process from training to results. There were no unexpected problems with the process. The shortcomings were normal, as a pilot have less resources and not always representative. The found problems will be fixed by changes in CAPI controls, a few changes in the questionnaire and longer training of the interviewers (40 days). It is difficult to evaluate the results from the pilot as the sample is not a true representative sample, but selected to show major differences between urban and rural areas. Some improvements have not been included, due to comparability with previous IOF. They will be subjects for the Modernizing Project.

Recommendations for INCAF/IOF 2014/2015

Strategic considerations at questionnaire design

The new questionnaires are primarily based on previous IOF that also followed the major international recommendations and best practices with the main exception that daily consumption was measured during only one week. Some other changes have been done to improve the quality, based on experiences from previous surveys, including INCAF. Most changes have been done in the labour force questionnaire. Priority users are National Accounts (NA) and Consumer Price Index (CPI) and IOF will closely follow international recommendations and classifications for those areas. For welfare and poverty analysis most changes are done to make analysis even more multi-dimensional.

Simplicity

Simplicity has not been a serious consideration. Quantities with very limit, if any, use will still be collected. Those data should at least be given less priority in editing and processing the data to avoid delays of the main results. Most problems in the pilot were related to the quantities. The IOF 2014/15 will collect more data than previous IOF, mainly due to users' demand, e.g. for more environment statistics. The third training test was mainly about labour force data. Almost all interviewers scored less here than in previous tests, indicating problems with the questionnaire or trainers rather than the trainees. The trainees are more like ordinary people, so if they don't understand it is a risk that the respondents will have problems even if the interviewers are trained more.

4x4 methodology

Each household will be measured all quarters and all types of weeks of the month, to avoid misclassification due to seasonal variation and monthly variation related to salary payments (contrary to the pilot report). All households in one EA will be measured the same week.

	1st week in a month	2nd week in a month	3rd week in a month	4th week in a month
1ST quarter	EA x	EA y	EA z	EA w
2nd quarter	EA w	EA x	EA y	EA z
3rd quarter	EA z	EA w	EA x	EA y
4th quarter	EA y	EA z	EA w	EA x

Full year

In the pilot semi-durables were measured for one month each quarter. This will be changed to 3 months. At the first visit the interviewer will ask for past 3 month and at the following quarterly visits about expenditures since last visit. Besides reducing the standard errors it will make the start of the periods more clear. To simplify, durable goods could be collected in the same way.

Partial non-response

If a household has moved since the listing of households, the new household will be used as the sample is actually is a sample of dwellings. Some households will move during the survey year. If no one is moving in, the household should be regarded as over-coverage and reduce the frame with one household. If a new household has moved in, the new household will be used as substitute. This is with the assumption that the new is not much different from the old. Sometimes the old household has moved because they have become better off, but the opposite will also occur. Socio-economic statistics will mostly be based on the first quarter, so later changes will have no impact. For expenditure, the impact will be limited as long as these movements are limited. For alternative handling of these households, they will have own codes. Following the old household would not be cost- and time efficient. The only practical alternative is to treat those households as non-response and substitute them with the average household, which will not avoid the risk of bias.

Panel rotation

No panel rotation will be used during the IOF-module. It is planned to be introduced in the next round of INCAF.

Sub-samples

Sub-samples should be used when smaller samples are sufficient, e.g. if monitoring subjective well-being. Sub-samples can also be used to test improvements for future modernisation.

Lump-sum

The lump-sum methodology has been proposed earlier but not been included in IOF. It gives an alternative to measure consumption continuously in a complete but simplified way. It should be tested on small sub-sample (500 households) during the second half of the survey. It can be a subject for the Modernisation Project. See annex 2 for a proposed questionnaire.

Short vs. long lists of items

IOF is aiming to collect all consumption. It is done by asking people of what they have bought since the last visit (supplemented by diary keeping), not asking item by item from a list. The size of the list should therefore have no impact on the results. IOF 2014/15 will have longer lists than previous IOF. CAPI will also simplify the use of the lists.

COICOP

The codes are much extended for clothing. For food outside home, distinction is not made by type of venue (restaurant, café/bar or canteen) but type of consumption (food, non-alcoholic beverages or alcohol). A separate questionnaire could be used for a small sub-sample to estimate more details.

The pilot had too many 999 999 as COICOP code. Recoding at INE is costly and time consuming and risks the quality as central staff may lack enough information. The item lists should be updated to cover most of these items, even in local languages. The training and supervision should also address this problem.

Individual questionnaires

Individual questionnaires will be used to better collect individual expenditure, particularly food outside home and transport expenditures. The interviewers should probe for these expenditure for each person even if they are not at home at the visit.

Out-layers

The pilot had a second-hand car for 30 000 000 Mt. CAPI should give a signal for potential outliers.

Data transactions

Data are transferred by Bluetooth from the interviewers to the supervisors and from the supervisors to both regional offices (as back-up) and to INE HQ as RAR-files. Handshaking and encryption is not used. This should be improved in the future.

Unsafe areas (Sofala) and coming election

Some areas may be too unsafe to visit. If it is more permanent local interviewers can be used, but otherwise those areas will be excluded. The documentation will declare that the survey is representative except for those areas.

Elections have been going on during previous household survey without implications. The interviewers are trained to be neutral and will already have visited all households before the election. Accidents outside control of the survey can always happen.

Poverty lines

Regional poverty line is much recommended as the food culture differs in such a tall country. People's competence to get most energy for the money is more relevant imposing the same food basket on everyone. Household poverty lines would even be better. INE is not responsible for the poverty lines and poverty analysis, but it is highly recommended that scales of equivalent is decided and used instead of "per capita" particularly when half of the population is under fifteen and children need much calories. Bigger households have also scale of economy.

Multidimensional welfare/poverty

Like previous IOF, the new survey will generate statistics in many areas relevant for welfare: employment, education, health, housing, services, possessions, etc. It is up to the analysts to decide if those data will be combined into a composite index. The main problem is the lack of weights. Composite indices are often disputed because different opinions on the weights.

As MPD is responsible for poverty analysis, INE should present some indicators on the whole welfare spectra, where poverty is part.

A better alternative than welfare index is to use a subjective well-being approach focusing on service performance and quality of life (including good governance, human rights and participation).

Statistics Sweden is using a well approved method to measure service performance in most parts of public sector and also the citizens' satisfaction with local governments and the living conditions in the community.

The method has also been used in many of Statistics Sweden's project countries and together with the Swedish Children's Ombudsman in six countries to track children's well-being as they perceive it themselves, different from many problem-oriented surveys of children. (See annex 3 for more information.

Tabulation and dissemination

The first quarter will be tabulated similarly as the first quarter of INCAF, i.e. focusing on labour force statistics. Possession of goods can be compared with previous IOF and INCAF. Access to water, electricity and hygiene and reading ability can also be analysed already on the first quarter. The food share of the total budget (Engel's law) is probably the most robust indicator on economic welfare and can probably be analysed on quarterly base and compared with earlier data from IOF and INCAF.

It is also important to use the first quarter to evaluate the quality and to improve processes, as well as planning the future reports in more details.

The full year will be tabulated similarly as the previous IOF, focusing on socio-economic and consumption statistics. Specific module, as Tourism will be tabulated by relevant departments while MPD will be responsible for poverty analysis.

The full year report should also compare with the previous IOF.

An important indicator on multidimensional welfare is the distribution, usually measured with the Gini coefficient.

The first quarter is expected to be disseminated in November 2014 and the full year in November/December 2015.

Risks

With many visits to the households, the data collection is not expected to be at risk. As in previous IOF the main risks will occur in data processing. A robust database is needed.

Database design

IOF 2008 was finally stored in a well designed database in MS Access, but the data processing was not integrated. In the previous processing (including INCAF and the recent pilot), data have been transferred from CSPro and manipulated in SPSS with a set of SPSS files without links and a huge amount of redundancies. These steps are not automatic (like queries in Access, so if there are some updates in the source data, all manipulations have to be repeated. There is also a clear risk that data will be changed in later processing without being changed in the basic data. As the files are distributed to different users, the end result can be contradicting statistics. Following steps are proposed:

1 Data entry files (CSPro) should be saved as back-up before any manipulations;

2 Relational database (Access or SQL) should be the core database with final data and latest updates;

3 Output database (Access or Excel) for easy cross-tabulations and decorations.

In a relational database (e.g. Access or SQL), data are proposed to be stored in the following tables.

	Table	Variables (columns)	Comments
1	EA data	EA Província Urbano/rural Sampling weight	One record per EA
2	Administrative data *Identificação*	ID (número de referência) District Name of head Address Telephone number	One record per household

		Coordinates (GPS) Contacts Controls	

3	Community data *Questionario comunitario*	IOF ID Date All variables: - variable id - result	One record per community
Questionário do agregado familiar			
4	Socio-demographic data	ID Person/*No. De Ordem* All variables in section 1-3	One record per person No names
5	Housing and household data	ID All variables in section 4-7	One record per household Expenditure in col. 18A-D(Y), 24A, 24C, 24E and 31a-b1 or b2
6	Anthropometrics	ID Person *No. De Ordem* All variables in section 8	One record per child No names
QUESTIONÁRIO DE EMPREGO			
7	Economic activity *Emprego*	ID Person/*No. De Ordem* All variables 1-21	One record per person
8	Income/*Receitas (seccao 2)* Transfers/*Transferências (secção 3)*	ID Person/*Pessoa* Item Value/*Valor*	One record per value
9	*XITIQUE RECEBIDO POR ANO*	ID Person/*Pessoa* Item Value/*Valor (12 months)*	One record per value Is not expenditure or consumption, but saving. Not part of traditional IOF Could be removed or saved only for specific uses.
10	Individual expenditure *Despesas individuais (secção 4)*	ID (Person) Item/COICOP Value	One record per value 7 days
11	Tourism/*Turismo (secção 5)*	ID Person All variables (3-19)	One record per person No names
12	Diary data *Despesas diarias*	ID COICOP Value/*Valor (col 5)* (Outlet/*Lugar (col 7)*)	One record per item/value No quantities Outlet = 5 for own produced
Despesas anuais, mensais e receitas			
13	Hosehold possessions *Posse de bens duraveis e despesas anuais*	ID COICOP/*Código* Possessed numbers/*Quantos possui*	One record per item 12 months

		Total value/*Valor total compra*	
14	Agricultural possessions	ID Número de ordem Quantos	One record per item
15	Insurances/*Seguros*	ID COICOP/*Código* Value/*Valor*	One record per item 12 months

16	Expenditure abroad *Despesas no estrangeiro*	ID Item Value/*Valor*	One record per item 12 months
17	Quarterly expenditure	ID COICOP Value/*Valor (col 5)* (Outlet/*Lugar (col 6)*)	One record per item/value
18	Gifts/*receitas*	ID All variables 1-13	One record per line Should be integrate into table 8

As long as CSPro is used data have to be transferred to the relational database. This can be done before or after data manipulation. Manipulations will derive some new tables.

After manipulations, following tables for analysis and final tabulations are recommended, excluding all identities (names, coordinates). Only variables expected to be used for standard/normal tabulation should be included. Specific needs should be treated separately. Quantities are only for specific use and should not be included in the main database.

|	Table	Variables (columns)	Manipulations
A	Sampling information	EA Provincia Urbano/rural Sampling weight	One record per EA Sampling weights updated for non-response (for each quarter). Separate set of weights for households participating all quarters.
B	Community data	IOF ID Date All variables: - variable id - result	One record per community
C	Individual data	ID Provincia (from A) Urbano/rural (from A) Sampling weight (from A) Person/*No. De Ordem* All variables in section 1-3, except names Anthropometrics (from tab. 6) Economic activity (from tab. 7)	One record per person Some variables (education, anthropometrics and labour data may be derived to more user friendly variables at this stage
D	Housing and household data	ID Provincia (from A) Urbano/rural (From A)	One record per household Housing variables should be filtered/derived to a few variables reflecting the quality of the house

		Sampling weight (From A) All variables in section 4-7, **except expenditure** Relevant characteristics of the head (from C) Household size/number of members Possessions (from 14-15)	Relevant characteristics are those expected to be used in standard tabulations

E	Income and transfers	ID Provincia (from A) Urbano/rural (From A) Sampling weight (From A) Person/*Pessoa* Relevant individual data (from C) Item Value/*Valor*	One record per value Some types of income and transfers could be aggregated if more user friendly Individual data relevant for standard tabulations of income particularly sex, level of education and employment status
F	Tourism/*Turismo (seccao 5)*	ID Provincia (from A) Urbano/rural (From A) Sampling weight (From A) Person/*Pessoa* Relevant individual data (from C) All variables (3-19)	One record per person No names
G	Macro-economics	ID COICOP Detiled item Weighted values	One record per item. All data weighted to annual values for all households in Mozambique. The macro-economic statistics for NA and CPI can be based on all respondents in each quarter, weighted with quarterly weights, or be based on households participating all quarters multiplied with a unified weight.
H	Household consumption	ID Provincia (from A) Urbano/rural (From A) Sampling weight (From A) Relevant household characteristics (from C) Adult equivalents Quintiles 1a Food expenditure 1b Own produced food (outlet=5) 2 Alcohol and tobacco 3 Clothing and footwear 4a Imputed rent 4b Housing expenditure 5 Furnishing 6 Health	One record per household All values weighted with sampling weights and to 12 months, i.e.: • Diary (from tab. 12) and individual data (from tab. 9) are multiplied by 13; • Quarterly data (from tab.17); • Monthly data health expenditure, rent and imputed rent[1] (from table 5); • Annual data (from tab. 13, 15, 16 and education expenditure in table 5) will have no time-weight. Adult equivalents could be derived from the data or the OECD adjusted scale can be used. If quintiles will be used, they must be based on household consumption per adult equivalent (contrary to previous IOF). There is no need to deflate the values, as the full year will be covered for each household.´ If the data are not equally spread between weeks

---

[1] Data for imputed rent can either be used as they are, or stratified. Stratification can be used to use some robust help variables representing the locality, size and quality of the house. The stratification in IOF 2008 may be too sophisticated.

		7 Transport	of the month, as planned, they may be calibrated so each type of week will have the same representation.
		8 Communication	
		9 Recreation	
		10 Education	COICOP codes are not needed in this table.
		11 Hotel and restaurants	
		12 Misc	
		Total	

Tabulation can easily be done with the pivot table function in Access or Excel. Access 2013 has no longer this function but can create automatic links to Excel for cross tabulations.

To create the socio-demographic tables for both households and individuals it is just to put (drag) the weights (representing the number of households) to the *cells* and set the *Autocalc* to *Sum*. Then different background variables can be dragged into the columns and rows to produce the different cross tables. In the picture below it is age in rows and sex in columns. Variables can easily be grouped and percentage tables produced by clicking *% Show As*. Most of the other modules like tourism, labour force, confidence, etc. can be tabulated in the same way.
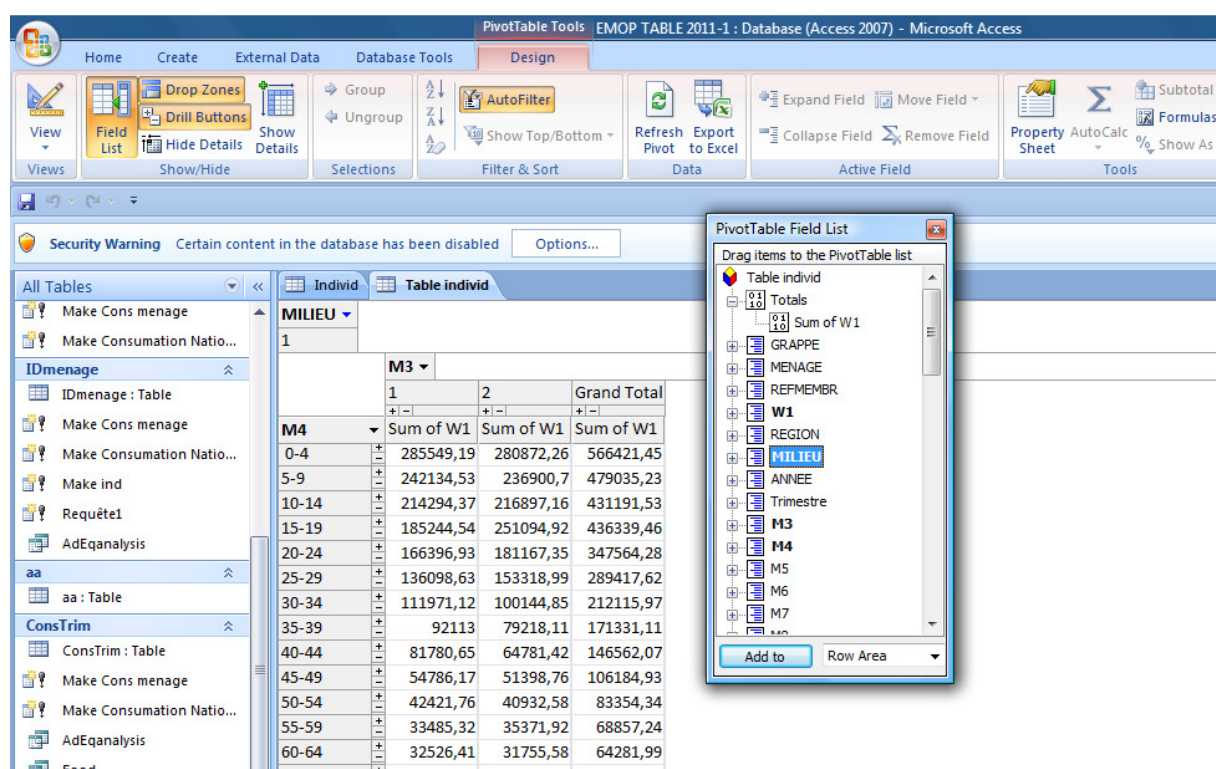
The economic data can be produced by bringing in and sum the weights into the cell area and also bring in and sum the consumption groups (1,food, 2 alcohol…, etc.) in the cell area. With the Formula function the economic data can be divided by the weights to tabulate household average consumption for different groups of households (as rows).

Table Consomation Menage

REGION	Households	1	2	3	4a	4b	5	6	7	8	9	10	11	12
1	176 016,18	360 263,00	7 720,29	30 292,52	7 723,55	7 720,29	26 615,35	20 715,16	43 256,38	32 834,53	11 248,19	6 422,63	3 237,70	13 096,30
2	230 040,16	290 327,26	10 833,96	34 033,93	5 159,46	10 833,96	17 695,20	19 453,41	70 980,02	48 505,18	11 106,27	11 572,59	385,60	31 760,08
3	250 870,35	247 172,35	2 924,20	17 821,38	7 930,47	2 924,20	21 180,65	16 657,95	39 876,60	35 765,08	6 717,09	4 759,14	903,89	30 547,63
4	224 444,38	231 114,18	3 418,79	17 949,27	4 913,87	3 418,79	18 564,07	16 345,75	36 516,13	21 867,75	6 105,90	3 832,78	418,85	14 539,44
5	200 196,34	275 220,05	3 668,59	18 687,60	3 612,20	3 668,59	21 377,03	29 639,39	48 103,08	26 566,74	6 977,28	3 703,63	250,53	17 050,37
6	96 067,50	265 463,15	2 140,54	24 720,71	916,99	2 140,54	12 204,92	18 211,24	12 370,74	15 797,74	11 292,78	263,33	167,17	7 641,69
7	80 302,61	287 706,90	4 346,09	22 903,21	6 007,57	4 346,09	16 188,15	16 157,82	50 753,46	30 721,97	5 041,83	10 471,78	273,28	11 053,55
8	12 879,24	330 830,90	9 590,42	40 596,35	8 039,03	9 590,42	12 342,67	16 172,67	34 237,70	15 890,75	9 789,23	737,37	443,51	6 481,24
9	222 613,35	285 418,02	8 259,40	37 110,20	22 714,01	8 259,40	38 016,14	26 799,19	154 549,27	81 890,25	15 483,94	19 004,38	2 495,03	22 111,10
Grand Total	1 493 430,11	278 273,68	5 760,77	25 712,57	8 097,07	5 760,77	22 504,87	20 840,50	61 523,86	39 208,95	9 407,79	7 830,70	1 090,55	20 475,62

Modernisation and future INCAF

INE has together with Scanstat started a modernisation project and INCAF/IOF has been selected as the first business case main object. Though many things are partly modernised with a continuous survey approach and introduction of CAPI/tablets there is still needs for further modernisation:

- The questionnaires could and should be simplified. All structural variables, particularly in labour force, should be rotating modules;
- The sampling design could be more efficient;
- The CAPI program currently based on CSPro should have better interface and support for the interview;
- A modern relational database should be designed for integrating and storing data;
- Metadata and documentation should be improved;
- Future dissemination should leave the concept of survey reports and not disseminate all kind of data quarterly only because they are available;
- A promising website for INCAF on INE's website is under development.

The future INCAF should have a regular module covering the informal sector.

Annex 1 - TERMS OF REFERENCE

for a short-term mission on
INCAF
the Continuous Multi-purpose Survey

9 - 20 June 2014

within the
AGREEMENT ON CONSULTING IN INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS
between INE and Scanstat.

Consultant:

Counterparts: Arão Balate and the INCAF team at INE

Background
INE has decided to modernize its system of household surveys by introducing a new Continuous and Multi-Propose Survey (INCAF) having a Core Module based on Labour Force and Expenditure where different modules like Informal sector, Tourism, Education etc. will be introduced quarterly. The INCAF module dominating INCAF 2014/2015 is the Household Budget Survey (Inquérito ao Orçamento Familiar, IOF).

Main reasons for the mission
The results from the INCAF/IOF pilot are soon available and there is a need for assistance and further advice on how to best go forward.

Objective
The overall objective of the mission is to assist INE in analysing the IOF-pilot using international experiences to optimize the implementation of the Continuous Multi-propose Survey IOF module 2014/2015.

Activities
This is a continuation of previous missions and the Consultant is responsible for the following activities:
1. Assist INE by giving recommendations for the new INCAF with income and expenditure module (IOF);
2. Support INE in the tabulation plan and its reports;
3. Share experience on conducting Continuous Multi-purpose Surveys as well at the preparation and presentation of the results ;
4. Assist with the tuning of the methodology by discussing with stakeholders regarding the use of the results;
5. The consultant will work closely with the Continuous Multi-propose Survey team at INE

Expected outputs
The consultant will prepare a written documentation of the above activities;

Beneficiaries of the mission
The mission will benefit INE and the whole National Statistical System of Mozambique including users of statistical information.

Tasks to be done by INE to facilitate the mission
- Elaborate ToR for the training
- Prepare and supply the consultant with necessary documents and information, like the new visual identity of INE
- Supply good communication conditions for the consultant.

Source of Funding
Project: MPD-2008-0006 – Inquérito Sobre Orçamento Familiar – IOF
        PAAO12 – 1.4.2 Inquérito Contínuo aos Agregados Familiares

Timing of the mission
Two weeks, as written above.

Place
The premises of INE in Maputo with possible allocations to the provinces.

Language
English.

Report
The consultant will prepare a short final report to be discussed with INE before ending assignment. Statistics Denmark as Lead Party will publish the final version on www.dst.dk/mozambique within 3+ weeks of the end of the mission. The structure of the report should be according to Scanstat format.

Approved by Arão Balate, INE/DCI

Day / / ....

Confirmed by Leia Gimo Macamo, Contract Manager for the INE – Scanstat Contract

Day / / ....

Annex 2 - A lump-sum approach

1 The lump-sum question should be included also in INCAF/IOF and developed:

- So lump-sum and diary data can be linked in the future,

- To make it possible to base the level of daily consumption on the lump-sum and the detailed distribution on the diary data,

- To make it possible to utilize 8 weeks daily consumption for each household, 7 days before the first visit and 7 days diary keeping.

Proposed questions:

Now I would like to ask if anyone in the household during the last week bought any food, drinks, hygienic products or other consumables (like paper, matches, cleaning and washing materials, etc., but not gasoline and fuels). **Don't include such things bought for agriculture or business.**

Did anyone buy those things last…	Yes	No	*If yes:* **For how much did the household buy in total?** *Mt*
..........day?	1	2	
..........day?	1	2	
..........day?	1	2	
..........day?	1	2	
..........day?	1	2	
..........day?	1	2	
..........day?	1	2	

In some shops it is possible to buy also other things, so I would like to know if any of the following items were included in the total spending you told me.

Did any of those spending include…	Yes	No	*If yes:* **What was the value of this?** *Mt*
...alcohol	1	2	
...tobacco	1	2	
...pet food	1	2	
...household utensils (matches, paper, detergents, etc.)	1	2	
…hygiene products (soap, toothpast, etc.)			
...clothing or footwear?	1	2	
...toys, games, sport goods?	1	2	
...magazines, books, stationary?	1	2	

Annex 3 - Welfare- as we perceive it ourselves

Typically, when constructing an index of welfare, a panel of experts decide what indicators to include and how they are to be weighted in order to form the index. But from the individual's point of view, the one who would be best suited to judge his own welfare is he himself. Here we present an alternative measure of welfare index based on questions to individuals.

There is a large and urgent need of comprehensive measurements of welfare. Traditionally most researchers use different ways of weighting indicators together to form a combined measurement. The choice of indicators and weighting system typically depends on the researcher. Of course this subjective choice might be open to criticism from other experts. Most surveys on welfare also focus on measuring the citizen's access to resources and service rendered by society. There is another way that has been used in a number of countries. It is a statistical methodology based on a structural model for creating welfare. This methodology is often used to measure customer perceived quality in the private and public sector, and also for staff satisfaction surveys.

The methodology applied to measuring welfare includes the entire process from access to service provided by the society (including information about service and possibility to affect the service) to willingness of using it. The methodology produces an index, which measures the individual's own perception of welfare in different respects, the confidence in society and also in other resources.

The totality and the components of the totality.

The method can be applied whenever you want to measure the levels of a "totality" and the components that make up the totality and also want to measure to what extent the different parts of the totality have an impact on the totality. Instead of as in traditional indexes first measure the components and then try to weigh them together into a comprehensive measure one measures the levels of the totality and the components and then compute the impacts of the components on the totality. Welfare, quality of life and its lower part, poverty, are entities that can be broken down into for example access to food, health, education, relations, safety, economical opportunities and so on. A representative sample of citizens is interviewed about how they perceive their living conditions both in general and specifically with respect to the different components. The calculation of the levels of the different components and the overall welfare as well as the calculations of the impacts of the components on the totality is done using a sophisticated statistical methodology. The methodology aims at finding a best fitting solution given reality as it is shown in the responses to the questionnaire.

Selecting indicators of welfare.

A first step is to gather interested stakeholders for choosing the components that are considered as relevant parts for explaining the overall perception of welfare. Each component that is part of the model has to be specified by a number of specific questions (indicators). Taken together the questions define the contents of the component. Representatives of different stakeholders like organizations, governmental authorities, experts and ordinary citizens are required to give suggestions on good and bad things in the living conditions. This is a sort of brainstorming in so called focus groups. Each suggestion is thoroughly discussed to arrive at a clear understanding of the concepts, which later will facilitate the interpretation of the results of the analysis. The final questionnaire will often consist of 2 to 4 pages with questions of the following type: How satisfied are you with your possibilities to influence politics? The answers are given using a 10-point scale, where 10 represents "completely satisfied" and 1 represents "not satisfied at all".

The collection of data is usually quick, simple and relatively cheap. Usually the survey can be added to existing national surveys as an add-on module.

An example from Jamaica.
Diagram 1 and diagram 2 show the results for men and women in Jamaica. The economical opportunities (the possibility to get work, credits to own enterprises and so on) gets the lowest score. The trust to politicians and authorities also gets low scores. The scores do not differ much for males and females, but the impacts are different. Food has a larger impact on welfare for women, while housing is more important for good welfare for men. A little surprising perhaps but the reason might be that in Jamaica the culture is such that men often have a loose connection to family and permanent living. The red staircase in the diagram can be seen as a rough help for giving priority to the components. The further below the staircase the more important is it to increase the satisfaction with the component in order to increase the perception of overall welfare.

High degree of explanation.
The correspondence between the index of perceived welfare and economical welfare, as measured by consumption, is strong. Citizens in Jamaica with a consumption level below the poverty line for Jamaica show an average index below 40 (transformed from the 10-point scale to a scale of 0 to 100). This can be compared with those having the highest level of consumption. They show an average score of more than 60.

The index of welfare is measured in the same way in different countries. Thus it is possible to compare the scores from different countries. By weighting the three indicators: satisfaction with the present living conditions as a whole, to what degree the expectations one year ago have come true and finally how well the present living conditions match the conditions of a ideal situation, the overall welfare index has been calculated for some supposedly rich and poor countries. The weights are automatically provided by the statistical method. A special version of the welfare measurement focuses on children and youths. Children in Sweden give a very high overall score compared to children in Kosovo. Grown up persons in Kosovo are not at all satisfied. The same goes for grown up persons in Serbia and in Saint Petersburg. In Jamaica the situation is better according to the welfare index.

Using larger datasets and comparisons to more "objective" indicators it could be possible to judge whether the scale is interpreted differently in different cultures. The degree of explanation is a measure of how well the model fitted describes the representative response set. It is often the case that about 60 to 70 percent of the variation in responses can be explained in small homogenous groups. For comparison, other similar surveys often have an explanation degree of around 20 percent.

The methodology is also very flexible. In order to try other perspectives it is possible to alter the components after the data has been collected. For example the traditional partition into sectors like health, education and so on can be used in one type of analysis, another partition in service offered by the society and individual participation and maybe a third based on the Maslow hierarchy of needs. The methodology can also be used as an interactive instrument for better understanding of the relations in a complex environment.

The methodology fulfills most reasonable demands for a comprehensive welfare index. The difference compared to other welfare indexes is that there is no need for expert judgments for weighting the components. The weighting is done by the representative sample of citizens in conjunction with statistical science.

Last but not least, the results obtained from the analysis offer an instrument for giving priority to official measures on different levels based on science. Because of the fact that the analysis

simultaneously measures both the levels of the components and their impact on the overall perceived welfare it can be used for giving priority to measures that are most efficient for increasing the citizens perception of their living conditions. The components that should be given priority are those that have low scores but strong impacts. Looking at the diagrams it is clearly seen how men and women value the different components both with respect to the present level and with respect to the impact the component has on the overall perceived welfare. Thus the methodology is clearly policy relevant for governments, a feature that has been used by several governments in Russia.

Diagram 1 Satisfaction with components of welfare and their impact on the overall welfare, for grown up men in Jamaica 2003
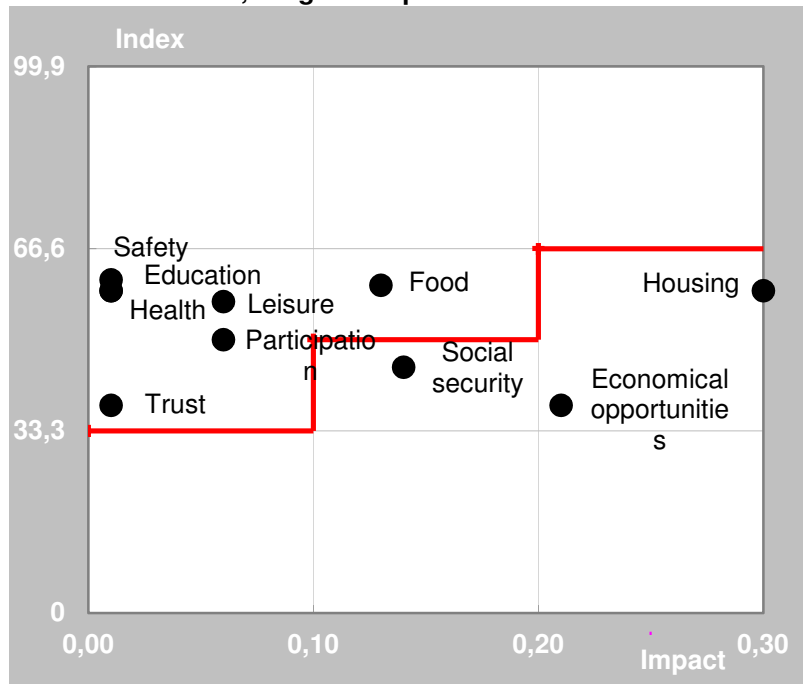

Diagram 2 Satisfaction with components of welfare and their impact on the overall welfare, for grown up women in Jamaica 2003
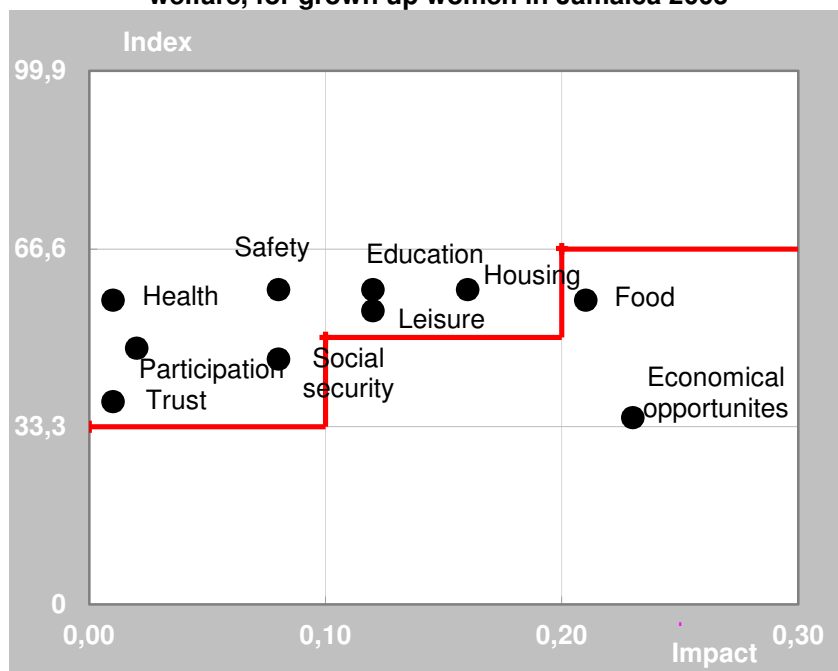
Table 1 Satisfaction with welfare by gender and level of consumption in Jamaica 2003

	Index
Men below the poverty line (40% lowest consumption)	40 + 3
Women below the poverty line (40% lowest consumption)	37 + 3
Men above the poverty line (40% middle consumption)	52 + 2
Women above the poverty line (40% middle consumption)	48 + 3
Men with highest consumption (20%)	60 + 3
Women with highest consumption (20%)	64 + 4

Table 2 Satisfaction with welfare for some countries/groups and degree of explanation

Country/Group	Index	Degree of explanation
Children (9-16) in Sweden, 2002	88	57 %
Children in Kosovo (9-15), 2003	60	63 %
Grown up in Kosovo, 2003	34	53 %
Grown up in Serbien, 2003	41	47 %
Grown up in Oman. 2000	46	46 %
Grown up in Jamaica, 2003	54	53 %
Grown up in St Petersburg, 2001	39	37 %
Grown up in St Petersburg, 2002	39	36%

Model explanation

For estimating the coefficients in the linear structural equation model we have chosen to use the PLS (Partial Least Squares) technique. This technique for solving structural equations is described for example in Fornell and Sha (1994), Lohmöller (1989) or Wold (1985). The motive for using PLS is that it provides a simple and robust technique. The core of the PLS algorithm proceeds in the following two steps:

(1) **Outside approximations**. Case-values of the latent variables are estimated as weighted means of the indicators according to

$$\hat{\xi} = \hat{w}_\xi x$$
$$\hat{\eta} = \hat{w}_\eta y$$

using weights $\hat{w}_\xi$ and w_η. The weights are the covariances between the latent variable and the indicators. The case-values represent the best predictors of the latent variables. For the formative case the regression coefficients between the manifests and the latent variables are used as weights. All variables are standardised.

(2) **Inside approximations.** Given case-values for the latent variables, improved values are obtained as weighted means of those latent variables that in terms of the inner structure are adjacent. Various schemes for weighting has been suggested see Fornell and Sha (1994)

After an initial estimation of the latent variables, the procedure iteratively switches between the inside and outside approximations until convergence. In a final step, the parameters of the inner structure, the measurement models are estimated. Each latent variable is determined by the inner structure and the measurement model. In each iteration, both equations are used to find an approximation of the latent variable. The estimated case-values will optimally fit into both equations. In the inner approximation the sum of the squared residuals from equation (1) is minimised ; in the outside approximation, the minimisation concerns the errors of the measurement equations.

The case-value estimates of the latent variables are transformed into a scale ranging from 1 to 100. They represent the individuals valuation of the respective component of the conceptual model. The estimates are obtained as the optimal (according to PLS) solution to the problem of adjusting the structure of the conceptual model to the empirical evidence given by the responses to the questions. The estimates of the coefficients in the structure represent the impact the different drivers have on the dependent variables. For example, an estimated coefficient of 0.8 from a driver such as "food" to total satisfaction means that if the satisfaction with "food" could be increased by one unit the increase of total satisfaction is expected to be 0.8 units. Thus the impacts can be interpreted as a numerical weighting of the latent variables according to the expected effect for changing the value of total satisfaction.