Albania - Living Standards Measurement Survey 2004 (Wave 3 Panel)

Institute of Statistics of Albania

Report generated on: September 9, 2016

Visit our data catalog at:

Sampling

Sampling Procedure

Panel sample, with LSMS 2002 and 2004

The APS 2004 collects information on 1,797 valid observations at household level and 7,476 at individual level. The sample of the second and third waves of the panel (APS) has been selected from the LSMS 2002 in order to be representative of Albanian households and individuals at national level. The LSMS 2002 differs from the APS 2003 and 2004 in that the former is designed to be representative at regional level (Mountain, Central, Coastal and Tirana) as well as for urban and rural domains, while the latter are for last domains only (urban and rural)

LSMS 2002 sample design

The LSMS is based on a probability sample of housing units (HUs) within the 16 strata of the sampling frame. It is divided in three regions: Coastal, Central, and Mountain Area. In addition, urban areas of Tirana are also considered as a separate region/stratum. The three regions are further stratified in major cities (the most important cities in the region), other urban (other cities in the region), and rural. The city of Tirana and its suburbs have been implicitly stratified to improve the efficiency of the sample design. Each stratum has been divided in Enumeration Area (EA), in accordance with the 2001 Census data, and each Primary Sampling Unit (PSU) selected with probabilities proportional to the number of occupied HUs in the EA. Every EA includes occupied and unoccupied HUs. Occupied rather than total units have been used because of the large amount of empty dwellings registered in the Census data.

The Housing Unit, defined as the space occupied by one household, is taken as sampling unit because is more permanent and easy to identify compared to the household. 10 EAs for each major city (75 for Tirana) and 65 EAs for each rural region -with the exception of the mountain area which is over-represented (75 EAs)- are selected. 8 households, plus 4 eventual substitutes, have been systematically selected in each EAs. As the LSMS consists of 450 EAs, total sample size is 3,600 households.

The sample is not self-weighted, hence to obtain correct estimates data need to be weighted. The weights, at household level, are included in the dataset ("weights" file). When working at individual level, household weights must be multiplied by household size.

APS 2003-2004 sample design

The panel component selected from the LSMS is designed to provide a nationally representative sample of households and individuals within Albania. It consists of roughly half of the households in the 2002 LSMS, interviewed both in 2003 and 2004. Contrarily to what done for the LSMS, no over-sampling in the Mountain Area has been performed for the panel survey.

The sample is designed to minimize the variability in households' selection probabilities. It insures national representativeness by matching the sample distribution across strata with the population distribution drawn from 2001 Census data. In Table 3 the ex-ante sampling scheme of the 2003-2004 APS is shown.

Compared to the LSMS design, statistical precision has improved. Under equal stratum population variances hypothesis, sample design effects are expected to be around 1.02, compared to the 1.28 of the LSMS sample. Moreover, further precision is obtained by keeping all 450 EAs of LSMS in the panel sample, thus reducing the eventual bias due to clustering because of new design.

Finally, the panel survey has a number of peculiar features that should be considered when using the data. The sample is designed to focus on individuals, who have been also traced when moving from the original household to a new one. This possibility represents the only way a household can enter the panel sample if it has not been already interviewed in the wave 1 (or in wave 2 for the APS 2004). If an original survey member (OSM) moves to a new household, his/her old and new household -and their members- are both included in the panel sample. Though a moved OSM will be present in the roster of both sampled households, he/she is a valid member only in the new one. In the old household he/she is taken into account as "moved away", hence not a valid member. This might generate some confusion.

Three modalities exist to classify an individual in the third wave. First, when he/she is an OSM, that is a respondent interviewed both in wave 1 and 2. Second, when he is a rejoiner from 2002, that is an OSM not interviewed in 2003 (i.e. because temporarily absent) who returns in 2004. Third, when he/she is a new member, whenever he/she is a newborn of an original household, a member joined by an OSM or a person who co-resides with an original survey household. So the APS is an indefinite life panel study, without replacement by drawing new sample units.

From wave 2, only individuals aged 15 years and over are considered valid members, hence eligible for the interview. Individuals moved out of Albania are not accounted as valid for this survey year, though they are still eligible for future waves.

Questionnaires

Overview

A single questionnaire on households has been used to collect information in the APS 2004. Contrary to the LSMS 2002 survey (see Basic Information Document, 2003), both in 2003 and 2004 the Diary for Household Consumption (the "booklet"), the Community questionnaire and the Price questionnaire were not repeated. The target is to collect a similar set of information (only data comparable across time is suitable for a longitudinal analysis) through a less lengthy and simpler questionnaire.

The household questionnaire has drawn heavily on the earlier APS 2003's -a reduced version of the 2002's-, but useful features have been added. Main changes are that data on credit have not been collected, the module on migration has been slightly reduced, while an additional section on remittances and a detailed module on social capital has been introduced. An additional module which collects data using pedometers on the distance (in kilometers) to the place of study and work of a sub-sample of interviewed households was introduced at an experimental level.

The choice of sections was aimed at matching as much as possible the specificity of Albania in terms of data needs, as driven by pressing policy questions. Their design (e.g. questions asked, their sequence, units and time-frame used) was adapted to fit the Albanian reality. Nevertheless, as consumption data are not available, the LSMS 2002 survey still represents the main household dataset for poverty analysis and evaluation.

Household membership is defined as being an actual resident or away from the household for less than six months (the exceptions being: the household head even though he might be away from the household for up to 11 months). Deceased individuals, lodgers, hired workers and servants are never considered household members. Individual in charge of answering the questionnaire is, usually, the most knowledgeable person about the specific matter. For the roster sheet for issued households, the one in charge of answering is the one designated by household members as the household head. If he/she is not available, a "principle respondent" is found. For the other questionnaire sections, identification codes of respondents indicate who provides information. In some modules where information is referred to individual, such as labor and health, each household member is asked to answer for him/herself. From wave 2, only individuals over 14 years old are eligible for interview.

As for the coding, ISCO (1988) and NACE codes are used for employment and industry activities, respectively.

Data Collection

Data Collection Dates

Start	End	Cycle
2004-05	2004-07	N/A

Data Collection Mode

Face-to-face [f2f]

DATA COLLECTION NOTES

Fieldwork

Implementation of the APS 2004 has been a cooperation among INSTAT, World Bank, and ISER (Institute for Social and Economic Research) of the University of Essex. Panel methodology has required acquisition of preliminary information, consisting of identification code (ID) of households, as well as names and ID of individuals to be interviewed. Information was pre-printed in the control form of the questionnaire (hhroster) which had to be completed from the interviewer through household information.

1,782 questionnaires with fed-forwarded information were issued and some left blanks for eventual split-off households. As the panel size of 2004 is the same of 2003's, 14 teams composed by 46 enumerators were involved in the fieldwork, as in 2003. The majority of enumerators had also participated in the 2002 and 2003 surveys.

Complexity of the panel survey method and of the questionnaire (see below) required a thorough manual questionnaires checking during fieldwork. This way, it was possible to re-interview the HU in case of errors and non-responses. Activity and profession coding took place at INSTAT, with the supervision of Word Bank.

Training and implementation

Staff training for fieldwork lasted a week, of which one day was spent on practice training with pedometer for the distance experiment, assisted by a World Bank consultant. Measuring distance has dealt with the use of pedometer whenever an individual walked to school or work. Operators had to teach the user how to perform the measurement and fill in the questionnaire. Results were reported in a separate form, and approximately 250 figures were reported overall.

Fieldwork followed three general rules:
a) split-off households were interviewed by the interviewers' team which had collected data from the original household (base family) if moved within same area;
b) information on the new household location was taken from the base family. Then, the questionnaire with the place information was send back to INSTAT which forwarded it to the team covering the district where the split-off individual had moved to carry out the interview
c) for households moved "out of scope" -so, impossible to be reached- the file was temporarily closed and an explanation note was added.

The fieldwork began on May 17, 2004. The interviewer workload was divided in two steps: 1) tracing and identifying sampled households and 2) carrying out the interviews. As previously mentioned, operator had to collect additional information for split-off and displaced households. Distances measurement by the pedometer was a new procedure, which increases the interviewers' workload. The average workload per operator was calculated in two questionnaires a day.

Fieldwork was completed within the expected deadline in early July 2004. About 1,812 questionnaires were completed.

Data Processing

Data Editing

A first data cleaning took place in Albania and implemented by INSTAT in collaboration with ISER and Government of Albania consultants. The cleaning process has involved following activities:
1. defining data checking routines and writing the syntax code of the cleaning programs;
2. generating lists of outliers and inconsistencies for each module to be checked against paper questionnaires;

During the first few days, data cleaning operators have been working on the Export Procedure of the Data Entry Program to check if data export succeeded and to finalize the English version of the dictionaries and error messages. Some changes were made to the Export Procedure due to a problem on the "Agriculturea2" file conversion and to the dictionary structure to check over correct labelling of exported data. The dictionary used during data entry was in Albanian language. So, an accurate comparison of the Albanian and English versions was done to ensure consistency (except for the labelling) between the two. This work was performed by using a freeware software called "Winmerge", which underlines all the differences between two text files.

Phase two has been devoted to update the Batch Edit (BE) procedure of the Data Entry Program, where a little correction was required to avoid some error messages incorrectly issued by the BE. Afterwards, the routine was applied to check all the errors, and a program in Access was run to associate PSU and Data-entry operator code to each questionnaire selected by the BE. Once obtained the procedure report, a pool of four people from INSTAT started to check all the reported errors and make the necessary adjustments. A copy of all original data in CSPRO software was made. During this work, some atypical circumstances were reported: sometimes errors or warning detected possible data-entry or interviewer problems. For these cases, no correction was made and the occurrence was highlighted in the report. Most of the problems reported by the BE were referred to the "distance that seemed to be inconsistent with the walking time" and "number of hours worked per week" higher than 70. All these situations were checked and corrected if differences between recorded values and paper values were found. At the end of these operations 10 problems (4%) were corrected out of 278 reported errors or warnings. Other 8 strange cases (3%) were underlined on the report.

The following steps were followed to check for questionnaire consistency. An SPSS program was written to check individual information present in the roster (Module 1) and coherence in the dwelling section (Module 2). No questionnaire was found to violate consistency in Module 1 and only one violating one of the dwelling rules.

Afterwards, a check of each qualitative variable value was carried out, by tabulating frequencies by variable and verifying if values were in their expected range. Any problem of this phase was reported, except for some "99" (meaning "not remember") values assigned to some date variables such as months or years.

Different criteria were used to check for outliers in quantitative variables. A variation of the classic interval around the mean was used for these cases. Since some very asymmetric distributions were found, a skewed interval around the median was adopted. This interval involved MAD (Median Absolute Deviation) and an asymmetric measure based on quartiles. For each module of the questionnaire a SPSS program was run to check the questionnaire consistency related to quantitative variables and to identify outliers. As in the case of Data-Entry checks, all these cases were verified and adjusted if a difference between recorded value and paper value was found. In suspected cases no change was introduced.

The third step was to check split-off households. The consistency dealt with verifying that all individuals not coded as present household members were considered as valid components only if they join a split-off household; otherwise they were considered "refused interview" or "impossible to be contacted". Only two cases of individuals without split-off household were found and the related corrections made.

Besides the preliminary check implemented right after the survey completion, additional controls were performed at a later stage. No major data entry errors were found, while some inconsistencies were highlighted and fixed. Furthermore, original files were reshaped to obtain individual or household-level dataset, as some initial data files were organized in a different way (e.g. by plot in the case of agriculturea or by income source in the socialassistance). Value labels for occupation and activity, whose coding was provided by INSTAT, were assigned to code-variables. A number of variables were created to better detect and trace households, and to enhance comparability across waves -e.g. cfinloc, which allows selecting households at their final destination, see below-. A careful check of each variable value was fulfilled, by tabulating frequencies and verifying if all values and codes were consistent. After the discrepancies were fixed, the general dissemination files were created.

Finally, an analysis of the differences in the sections' content between the two waves of APS was performed. This may be useful for analysts aiming at using the longitudinal panel of the three waves, if used in conjunction with the

Variable_reconciliation LSMS_PANEL_final document of wave 2. Codebooks for the individual and household-level files were created and are part of the documentation.

Other Processing

Linking files within wave 1

Data files from the household questionnaires can be linked by using the identifying variable for each household (hhid, labeled 'household identifier'). This is a three to five digit code, where the last two digits always represent the household number (from 01 to 14), and the first digits represent the PSU number (from 1 to 450). This variable is computed by linking the PSU (psu) identificator, and the sampled household progressive number within the PSU (hh, labeled "household ID"). For instance, household 4 in PSU 120 will be labeled as hhid=12004 (i.e. by combining PSU 120 with hh number 04). Individual level files also include a variable indicating the person whom information is referred to. The name of this variable varies across files, but it is usually associated with the label "ID Code".

Linking files within waves 2 and 3

Household-level observations can be matched by the household identifiers bhid and chid, for wave 2 and 3 respectively. Individual-level observations can be matched through b1_q01 and m1_q01. [Note: It is worth noting that m1_q01 is equal to the personal identification code (pid) across wave.] The difference between bpno and bpan_num (and, hence, between cpno and cpan_num) is that the former is created for all individuals in the roster, even if they are not valid members, while the latter has been created for valid household members who have compiled the questionnaire. To merge individual-level file w2_ind_all and w3_ind_all with household-level files, e.g. w2_hh_all and w3_hh_all, the user is required to use bhid within wave 2 and chid within wave 3.

Link observations across waves

As already mentioned, the aim of the Albanian Panel Survey is to collect information on the same analysis units across time. To this goal, a 9-digit personal identifier pid has been created, which is constant across time for each individual. So, the pid can be used to merge individuallevel observations across waves. Instead, following households across time is not as simple, because household identifiers (hhid, bhid and chid) are not constant over time. The variable hhid has been added to the 2003 wave and can be used to merge household information between wave 1 and 2. The variable m0_w2_hh, equal to the household identifier (bhid) in wave 2, has been included in the metadata file of wave 3 and it can be used to match information across the last two waves. Hence, through the above variables it is possible to follow a household in the entire panel.

Data Appraisal

No content available

Related Materials

Questionnaires

Albania Panel Survey - Wave 3 - 2004

Title	Albania Panel Survey - Wave 3 - 2004
Author(s)	Institute of Statistics of Albania
Country	Albania
Description	Household Questionnaire: This questionnaire is an essential tool for using the household and individual level data. It contains all of the questions included in the data and the codes for the responses and skip patterns where applicable.
Filename	alb04hhq1.pdf

LSMS 2004: Other Modules Household Questionnaire

Title	LSMS 2004: Other Modules Household Questionnaire
Author(s)	Institute of Statistics of Albania
Country	Albania
Description	Household Questionnaire: This questionnaire is an essential tool for using the household and individual level data. It contains all of the questions included in the data and the codes for the responses and skip patterns where applicable.
Filename	alb04hhq2.pdf

Technical documents

Basic Documentation: Albanian Panel Survey 2004

Title	Basic Documentation: Albanian Panel Survey 2004
Author(s)	Albania Institute of Statistics (INSTAT)
Date	2005-03-01
Country	Albania
Description	Describes the design of the survey and its coverage for potential users and provides general information about the general characteristics of the survey.
Filename	BasicInformationDocument_2004.pdf

Household Data Codebook

Title	Household Data Codebook
Country	Albania
Description	This document shows all variables and codes found in each of the data sets collected at the household level.
Filename	Codebook_w3_hh_all.pdf

Individual Data Codebook

Title	Individual Data Codebook
Country	Albania
Description	This document shows all variables and codes found in each of the data sets collected at the individual level.

Filename   Codebook_w3_ind_all.pdf