Basic Documentation

Albania: Employment and Welfare Survey
(August - November 1996)

Poverty and Human Resource Division
The World Bank

July 1998

Updated November 2003

ACRONYMS

MOLSP	Ministry of Labor and Social Protection
NE	Ndhime Ekonomica

Table of Contents

List of Tables

List of Boxes

This document was prepared by Carlo del Ninno (DECRG) with the assistance of Harold Alderman (DECRG).
Revisions were made by Diane Steele, Tilahun Temesgen, and Margaret Grosh (DECRG).

1. Introduction

The original objective for collecting a detailed household data set in Albania was to provide the data to support the research effort *Decentralizing Safety Nets: Community Choices and their Impact on Households*.[1] One of the main components of the research program was the evaluation of the distributional impact of the efforts of decentralizing the social assistance program, the Ndhime Ekonomica (NE), at the household level.

The Government started the social assistance program to provide targeted income support based on family earnings. Families applied to the Ministry of Labor and Social Protection (MOLSP) and provided information on land holding and income earnings. The amount of resources allocated to the communities was regulated and the initial level was later reduced by a fixed percentage. In March 1995, 20 percent of the population lived in families that received NE assistance. In October 1995 a new law was implemented which gave the NE more features of a block grant. Local authorities were allowed to keep 50 percent of the difference between the grant and the entitlements to be used for community projects. Concurrent with the development of the NE, a rural poverty alleviation project was expanded from a pilot to full implementation. This project has a public works component for the restoration of rural infrastructure.

The data collected for the Employment and Welfare Survey include information on household composition, education, current and past employment, level of expenditure, health outcomes, and other important household characteristics such as assets and quality of housing. In essence the data collected include all dimensions of household characteristics which makes it important in analyzing the original goals of the research project as well as to perform many other welfare and poverty studies.

The data collected cover approximately 1,500 households in rural and urban areas, excluding Tirana. Tirana was not included because it had been the focus of another household survey in 1994. The sample was designed to maximize the inclusion of poor areas to increase the number of program participants. The actual data collection took place between August and November 1996.

This document is organized as follows: Section 2 briefly describes the survey instrument. In Section 3, the sample design and the weighting methodology are discussed. Section 4 describes the protocols for field work. The data entry and data management are discussed in Section 5. In Sections 6 and 7 the original and supplemental data files are outlined. Notes to potential users are presented in Section 8.

[1] Decentralizing Safety Nets: Community Choices and their Impact on Households. RPO680-98. For more information on this and other research projects, see the World Bank's Development Research Web Site at: http://www.worldbank.org/html/prddr/prdhome/index.htm

2. Survey Instruments - Questionnaire

2.1 *Summary description of the household questionnaire*

The household questionnaire design follows the general principle of a Living Standards Measurement Study (LSMS) survey.[2] It includes variables necessary to describe and model several dimensions of the household in detail. The employment section contains information on current employment and a section on job history. The expenditure section contains detailed food and non-food expenditures that can be used to calculate levels of welfare. The sections on employment status and history, and social assistance contain specific details to respond to the needs of the original research purpose.

The questionnaire is organized in 12 main sections and several subsections. A complete copy of the questionnaire is available upon request (see Appendix D). A brief description of the contents of the questionnaire organized by the main sections is presented in Box 1.

2.2 *Special characteristics of the questionnaire*

One of the goals of the survey was to obtain data on the changes that have taken place since the first democratic election that took place in 1991. These include information on household members that left the household and the type of jobs held, the acquisition of land, livestock and other assets. For employment and social assistance there was an attempt to reconstruct a timeline of employment and participation in social assistance since it's inception in 1992. Therefore in several sections of the questionnaire the information is requested for the current and each of the past three years.

Detailed nutrition and anthropometric data were not collected. An agreement had been made with UNICEF that they would follow the same households and collect those data a few months after the initial survey. Unfortunately, the political situation prevented this from happening.

3. Sample Design

3.1 *General description and sample size*

The sampling frame includes all rural and the urban areas of Albania except Tirana (see Introduction). The sample was drawn using a multi-stage stratified sampling procedure separately in the rural areas and in the urban areas.

[2] For more information on the LSMS visit our web site at: http://www.worldbank.org/lsms/lsmshome.html

Box 1. Summary Description of the Content of the Questionnaire

1.	Household information	Sec 1.1 contains the roster which includes gender, age, civil status,etc. In addition it asks if the individual moved to the current location after the elections of 1991. Sec 1.2 contains information on former household members that left the households after the 1991 election.
2.	Education	Sec 2. Is limited to two questions for all individuals age 6 months and older: a) highest level of formal schooling completed and b) if the respondent is currently in school.
3.	Status and history of employment, job search, training and public works	The employment section was administered to all household members age 16 and over. In Sec 3.1 there are questions relative to current labor participation, the main type of work and looking for work if unemployed. Sec 3.2 contains questions relative to the respondent's job search strategy and the attitude towards accepting a job (willingness to relocate and minimum wage). Sec 3.3 contains information on employment before the 1991 elections. Sec 3.4 reports participation in public works and training.
4	Current main and second job	Sec 4.1 and 4.2 contain information on primary and secondary jobs, respectively: Type of job, industry, time allocated, type of contract, salary and benefits.
5	Non-agricultural self employment, business assets and durables	Sec 5.1 contains information on time spent and revenue received for self-employment as priimary or secondary work. Sec 5.2 contains number, value, estimated age and ownership of business assets and durables and the transactions for equipment and value of inventory during the past 4 years
6	Agricultural activity, agricultural land, agriculture assets and livestock	Sec 6.1 reports the number of weeks worked in agricultural activities during the past year and the hours worked last week. Sec 6.2 reports details on access (for each of the past 4 years) and type and acquisitions of agricultural land (orchard, pastures and crop-land). The information on agricultural assets in Sec 6.3 includes quantity , value and ownership. Sec. 6.4 reports type of livestock owned and the level of the stock at different points during the past 3 years.
7	Remittances	Remittances sent or received, Sec 7.1 and 7.2 respectively, are the mirror image of each other. The amounts exchanged are relative to the last 6 months and in each of the past three years

8	Social assistance, insurance	Sec 8.1 collects information on specific benefits received in 1993, 94, and 95 including amount and number of months received. Sec 8.2 collects informatino on benefits that were applied for, but refused.
9	Housing, ownership of real estate assets, Household furniture and durable goods	Sec 9.1 collects information on current housing, including type of building, ownership status, rent paid, and amounts paid for servicees. Sec 9.2 provides information on real estate assets, including type of property, when and how it was acquired and the current resal value. Sec 9.3 collects information on household furniture and durable goods not used for business.
10	Food expenditure and consumption	The food section contains 53 food items in 9 food groups. For each item quantities consumed during the last month, from purchases, own production, and received from other sources are listed along with the purchase value (if quantities are not known), and their current prices.
11	Non-food spending	Sec 11.1 collects information on personal items and regular transportation costs during the last month, and Sec 11.2 collects informatino on personal items, construction materials, clothing, and health care items for the past year.
12	Health status	This information collects information on type, duration and treatment for chronic illnesses (Sec 12.1), and type, absences caused by, and treatment for acute illnesses (Sec 12.2).

In the first stage, formal administrative units were selected with probability proportional to the number of individuals receiving economic assistance (NE). The administrative units are communes in rural areas and bashki in urban areas. The areas with more participants were over-sampled to have a larger pool of NE recipients. The behavior of recipients can be compared to non-participants in the program because they were drawn randomly. In the second stage, households were selected with equal probability as described below.

Selection of rural households

In rural areas the following selection process was used. In the first stage, 50 communes were selected from the total of 314 communes in Albania. To do this, the communes were first arranged randomly and then they were selected with probability proportional to the number of estimated recipients, using the cumulative list of number of recipients to give higher probability of selection to the communes that have a larger number of recipients.

In the second stage, a target of 29 households was defined and selected with systematic random sampling using the civil registry lists available in the communes. These lists are kept up to date and include all of the households that have formal residence in any of the villages that form the commune. The actual selection of households was based on the interval that was

calculated using the latest data from the MOLSP and a random starting number. The interviewers were instructed to go down the list and select as many households as they could until the end of the list. In practice, the actual number of households to be interviewed was based on the number of households reported in the civil registry. Note that the number of households reported in the registry could be different from the latest data reported at the MOLSP due to possible imperfections of the lists. When more households were found in the registry than anticipated by the MOLSP list, more than 29 households were selected to be interviewed, and when fewer households were found than anticipated on the list fewer were selected. Since the lists are organized by villages, this procedure assured that the number of households selected in each of the villages in a commune reflects the relative population in each village. With the exception of very few cases in which the number of households selected was lower than 29 (for example, Diber with 16), the number of households to be interviewed was close to the expected number.

Selection of urban households

A slightly different procedure was followed for the selection of households in urban areas. In the first stage 8 out of 48 bashkies were selected with probability proportional to the number of recipients of the NE program. In the second stage a fixed number of 60 households was selected in each bashki with systematic random sampling. Because civil registries in urban areas were not maintained up to date, it was decided not to use them. Instead all households in the eight bashkies were physically listed and then a fixed number of households were selected from these newly created lists. In addition, to reduce the cost of listing, the larger bashkies were segmented into subsections that were selected randomly for the analysis. In this case no assumptions can be made about the urban population being different from the latest official data.

Table 1. Sampling Frame by Location

	Rural areas	**Urban areas**
Number of Communes / Bashkies	314	48
Population	2,004,082	926,563
Approximate N of Hhs	496,060	231,269
Approximate Hh size	4.04	4.01
Percentages of Recipients	18.93%	18.40%
Number of expected interviews	1450	480

The selection intervals and all other information relevant to the selection process are depicted in Appendix B. These tables refer to the selection process in the urban and rural areas respectively.

3.2 *Non-response and replacement procedure*

The original sample was drawn without replacement and interviewers were instructed not to replace households that were not found or that refused to participate in the study. One of the reasons for selecting the sample without replacement was to fill a specific request on the part of MOLSP to get a rough estimate of the migration from rural areas. Apart from those non-

responses for households that were not found at their residence and that were most likely to be migrants, the rates of non-response and refusals have been kept to a minimum. The reasons for non-responding households for both urban and rural areas are reported in Table 2.

As expected the pattern of non-response is quite different between urban and rural areas. In urban areas the listing operation was carried out shortly before the survey. Therefore, the bulk of non-responses is caused by actual refusals and other non-specified reasons.

Table 2. Number of Completed/Incomplete Interviews and Non-Response by Type

	Urban		Rural		Total	
Not interviewed	**Number**	**Percent**	**Number**	**Percent**	**Number**	**Percent**
No info	12	17.39	39	11.40	51	12.41
Migrated	1	1.45	232	67.84	233	56.69
Refusal	29	42.03	11	3.21	40	9.73
Not completed	27	39.13	60	17.55	87	21.17
Total not interviewed / Not completed	69	100.00	342	100.00	411	100.00
Total Planned	480		1433		1913	
Total Completed	411		1091		1502	

In rural areas, non-responses caused by migration were substantial, about 12 percent of the overall planned sample and two-thirds of rural non-response. This confirms the existence of a current migration stream from the rural areas into the urban areas and abroad. This phenomena may not be apparent in the civil lists because in remote rural areas dwellings cannot be rented out or sold. It is also possible that the migration is just temporary and the households prefer to maintain a formal residence in rural areas to have a place to go back to in the future.

3.3 *Probability of selection and weighting*

The probability of selection and the resulting sampling weights were calculated separately for the rural and urban areas. In general, the probability of selection is:

$$P(\alpha\beta) = a * \frac{R\alpha}{\sum R\alpha} * \frac{b}{M\alpha}$$

Where: a = Number of communes / bashkies
b = Number of expected households
R_α = Number of recipients in cluster α
M_α = Estimated number of households in cluster α

Using the corresponding weights (the inverse of the probability of selection) for each area (for both the recipient and non-recipient households), the correct population projections can be

obtained for both rural and urban areas. In other words, when the data are weighted the results are representative of rural and urban Albania with the exception of Tirana which was excluded from the study. When weights are used, the resulting estimate of the number of recipients found in the field are correct.

Since non-refusing households were not substituted, the probability of selection was adjusted to take into account the probability of non-response, i.e. both probabilities of selection were multiplied by the ratio of completed questionnaires over total attempted questionnaires.

$$P(\alpha\beta) = a * \frac{R\alpha}{\sum R\alpha} * \frac{b}{M\alpha} * \frac{Hh\,Re\,sp}{HhEnum}$$

Where: Hh Resp = Number of households actually interviewed
HhEnum = Number of households enumerated (attempted to be interviewed)

In this case the actual calculation of the probabilities of selection in rural and urban areas is slightly different. In rural areas the interval of selection was fixed, not the number of households to be collected, therefore the expected take is different from the actual number of households to be interviewed (enumerated). In urban areas the number of households to be interviewed was fixed and b is equal to HhEnum.

In conclusion using the urban and rural weights calculated as shown above is believed to provide good estimates of the number of households in Albania. We did find a smaller number of households than expected in rural areas, but a larger mean household size. Part of the reason for the undercount was caused by the extensive migration from rural areas and an adjustment was made to the weights. Details pertaining to this adjustment are reported in Section 8.2.

4. Organization and Fieldwork Procedures

The data collection in the field was contracted by the MOLSP to a local consulting company. The contract was funded by the monitoring and evaluation funds for the Social Safety Project funded by the World Bank

The main difficulties for the field work and implementation of the survey were caused by transportation problems, especially in rural areas. Although Albania is a fairly small country, road transportation is particularly difficult and several villages in rural areas are connected to the main road only by off road mountain passes and hiking trails.

4.1 *Design, translation, testing*

The questionnaire was designed in collaboration with individuals from the MOLSP in Albania and at the World Bank. The first version of the questionnaire was tested in the field in May 1996 and the final version was prepared in July/August 1996.

The questionnaire was translated into Albanian. The quality of the translation was

verified by the collaborators at the MOLSP at the time of the field test when the answers coming from the field were discussed.

4.2 *Structure of the interview*

The administration of the questionnaire took about one hour (60 minutes on average in rural areas and about 70 minutes in urban areas). The questionnaire was administered in one single visit, depending on the households ability to provide all the answers during that first visit. Because of the difficulty of transport in rural areas and the feasibility of conducting the whole interview in a single visit, the one visit strategy was the most efficient in Albania.

4.3 *Structure of survey teams*

A total of five survey teams were used for the data collection. Each team consisted of a field supervisor, three interviewers and a driver. In total seven supervisors and 15 interviewers worked on the survey. Each interviewer completed on average a total of 100 questionnaires, ranging from a minimum of 40 to a maximum of 182 households.

The quality of work provided by the interviewers was very high. Most of the interviewers had a college degree and most of them had been, or still were, school teachers.

4.4 *Training*

The training took place at two points in time: prior to the initial field test, and in August 1996 just before field operations started. The training included detailed discussions of the questionnaire over a period of two days, some practice interviews under the supervision of the trainers and discussions of the completed practice interviews. Given the small number of teams (only five) and the large number of interviews completed by each interviewer, the quality of data collection was good.

4.5 *Publicity and remuneration to the households*

The fieldwork was conducted using a very low key approach. The survey manager visited the communes and discussed the survey and its purpose with the local Government officials. They obtained the registration list for all households residing in the various villages and used the list to select the households to be interviewed.

A small present was also given to the households that participated in the survey. The exact present given varied from time to time. Sometimes it was an item for school children, like a book bag or coloring pencils. Other times it was a household item like a clock.

The reception given to the interviewers was very good especially in rural areas. The participation of the households was excellent. This attitude was reflected in the low number of actual refusals and minimal item non-response.

5. Data Entry Management

5.1 *Data entry program and verification*

The data entry program was designed using IMPS, a data entry package developed by the US Census Bureau. The program was designed in such a way to follow the same layout as the questionnaire and included three types of data checks: a) range checks; b) intra-record checks to verify inconsistencies pertinent to a particular section of the questionnaire; and c) inter-record checks to determine inconsistencies between the different sections of the questionnaire.

Some detailed checks were performed to verify the wage level, food prices and also if the per capita quantities of food were inside defined acceptable ranges. The program for checking the first two types of errors were performed simultaneously as the data were entered in the computer. The inter-record checks were performed for one or more households at one time, after the data were entered in the computers. The original IMPS dictionary of variables and the program for error checking is available upon request (see Appendix D).

5.2 *Data entry operations*

The screen layouts of the data entry were translated in Albanian and the error messages into Italian to facilitate the task of data entry operators.[3] The data were entered in Tirana, where all operations were coordinated, by seven data entry operators that worked two to three shifts a day. The data entered using IMPS can be stored in files that contain the data relative to one or more household questionnaires. In this case it was decided to group households by the communes/bashkies where the data were collected. This facilitated the management of the data and the checking of errors for the inter-record checks. The error report relative to the whole commune was prepared and given to the supervisor that worked in that area, but revisits to households were not done on a routine base.

5.3 *Organization of data for analysis*

Once the data had been collected and checked with the data entry program the data files for the analysis were prepared. A special custom program, designed in Microsoft Access, was prepared to read the original IMPS files, to rearrange them according to the sections of the questionnaire, to add additional labeling information and to create the programs used by the statistical packages to import the data. While it is convenient to store data at household level when they are entered (to facilitate running inter-record checks), it is difficult to analyze it in that format. The 58 data files that resulted from the key entry of data from the 50 communes and the 8 bashkies were re-grouped into 35 files organized according to the sections of the questionnaire (see Box 2).

5.4 *Preliminary data cleaning*

[3] This mix of languages was at the request of the Albanian data entry operators. Italian is widely spoken in Albania.

Besides the original checking and cleaning that was done in Tirana at the time of data entry, some additional cleaning was done during the analyses. Any data that did not look right were sent back to Tirana for verification where the consulting company verified the information against the original questionnaires.

6. Organization of Electronic Data Files - Original data files

6.1 *Description*

The electronic data files derived from the survey and prepared for analysis include 35 hierarchical data files. The names of the files follow the same pattern and structure as the questionnaire. Some of the files are organized at the household level, some of them at the individual level, and some of them at the food commodity level, depending on the type of information included in the file.

Each file contains identification variables that allow for the merging and matching of the information to create new files that contain the variables needed for the analysis. A detailed list of the 35 original data files is in Box 2.

6.2 *Documentation, codebook and summary statistics*

The questionnaire contains most of the information needed to interpret the data. The questions have been laid out clearly and the interviewers were instructed to follow the questions literally. Most instructions are printed in the respective sections. Most of the codes needed for the responses are included in the box relative to the question. In a few cases where the codes included a large number of responses or the responses were lengthy, they are reported in a box on the same page as the questions being asked.

There has been an intensive use of skip patterns to facilitate data collection and minimize the time spent filling in the questionnaire. Skip patterns are represented by an arrow followed by the number which refers to the next question to be asked or to the next section (e.g., →8). In all the cases where a skip pattern applies, the data in the skipped questions will appear as missing. Truly missing values refer only to the questions that were not supposed to be skipped, but that received no answers.

Box 2. Original Data Files

Num.	*File Name*	*Description*
1	SEC00	sec00: cover page
2	SEC011	sec011: household roster
3	SEC012	sec012: former household members
4	SEC02	sec02: education
5	SEC031	sec031: status of employment
6	SEC032	sec032: those currently looking for a job
7	SEC033	sec033: history of employment
8	SEC034A	sec034a: training and public works
9	SEC034B	sec034b: training and public works
10	SEC041	sec041: main job – current
11	SEC042	sec042: second job – current
12	SEC051	sec051: non-agricultural self employment
13	SEC052A	sec052a: business assets and durables
14	SEC052B	sec052b: business assets and durables (cont.)
15	SEC061	sec061: agricultural activity
16	SEC062A	sec062a: access to agricultural land
17	SEC062B	sec062b: agricultural land
18	SEC062C	sec062c: agricultural land (cont.)
19	SEC063	sec063: agriculture -- assets
20	SEC064	sec064: agriculture -- livestock
21	SEC071A	sec071a: remittances received?
22	SEC071B	sec071b: remittances received
23	SEC072A	sec072a: remittances sent?
24	SEC072B	sec072b: remittances sent
25	SEC081	sec081: social assistance
26	SEC082	sec082: benefits availability
27	SEC091	sec091: housing
28	SEC092A	sec092a: own real estate?
29	SEC092B	sec092b: real estate assets
30	SEC093	sec093: household furniture and durables
31	SEC10	sec10: food expenditure and consumption
32	SEC111	sec111: regular non-food spending
33	SEC112	sec112: occasional non-food spending
34	SEC121	sec121: health status
35	SEC122	sec122: health status - acute illness

Simple descriptions of files, variables and code labels, along with simple summary statistics, are provided in text and Microsoft Word format. Summary statistics are very useful to get a feeling of the data before they are used for analysis. They are also very important for use in verifying that the data received are complete and have not been modified. Sometimes a small translation mistake from one format to another might change the nature of the data.

6.3 *Missing data and other special codess*

There are several questions that have been coded as ‘yes’ and ‘no’. The codes used this time have been 0 and 1 where Yes=1 and No= 0.

Missing values have been left blank on the forms and they are treated as "." in most statistical packages. Refer to the statistical package used to get more details about their treatment. For example, in STATA they are not used to calculate sample statistics and they are assumed to be the largest numbers in the data set. In a very few cases special codes "99" or "98" have been used to highlight special situations. These are clearly marked in the questionnaire.

6.4 *Merging data from different data sets*

Each household can be uniquely identified using the household identification variable *hhnum*. This is a variable made of 6 digits. The first two digits refer to a sequential code identifying the commune or the bashki. The following two refer to the code of a village of the commune or a zone in the bashki. The last two digits refer to the sequential number of the household. Information regarding the specific communes/bashkies and villages can be found in the Tables B1 and B2 in Appendix B. It is also included as the variables *bashki* (code for the commune in rural areas and the bashki in urban areas), *sector* (Village/neighborhood) and *hhold* (Household number) in the file COREHH.

The *hhnum* is the only identifier for household level files. Individual level files have an individual code in addition to the household code: *pcode* (individual id code). Of course all the individuals in the same household share the same household code. This permits analysts to clearly identify the information relative to each individual in the data set. Similarly, other files of a different level of aggregation have additional identifiers. For example the food expenditure files have a unique code for each commodity that has been consumed by the household: *fooditem* (code of food).

This identification method assures that the data are stored in the most efficient way and that data in different files can be easily combined in analyses to compare the existing variables and to create additional ones. Household level files can be joined together to combine variables from different files using the variable *hhnum*. Similarly, individual level files can be merged together using variables *hhnum* and *pcode*. It is also possible to add household level information to an individual level file or a community level file. The only caveat is to be careful about the keys that are used to sort and merge the data and make sure that the resulting file contains the data for the same individual, household etc.

7. Organization of Electronic Data Files - Supplemental and Constructed Data Files

7.1 *Description*

There are three additional files that are part of the core data set. These files include information that is needed in order to be able to process the data properly. They include a file with the information about the sample and the weights, a cross cutting file with information on location and other often used variables and a file that contains the calculated measure of welfare. These files are reported in Box 3 and the summary statistics are available in Appendix D.

The file WEIGHTS contains 58 observations relative to the basic data and the weights for the 50 communes and the 8 bashkies included in the sample. The key variable is *bashki* and identifies the commune or the bashki. It also matches the first two digits of the household identifier variable *hhnum*.

Box 3. Additional Data Files

Num.	*File Name*	*Description*
1	WEIGHTS	Weighting and sample selection file
2	COREHH	Basic Hh Info file with weights
3	EXPAGG	Total Aggregate expenditure

The file COREHH is one of the most useful files to create interesting tables for any aspect of the data. This is a household level file that contains variables relative to the household size, the household composition, location variables (*bashki*), type of settlement, regional categories, date of the interview and the weight variables. The variable *weight* is the original weight variable. This file can be easily merged with any other file using the household identification variable *hhnum*.

The file EXPAGG contains a measure of total welfare, calculated as total aggregated expenditure, that is made available to interested users. A detailed explanation of the methodology used to calculate total expenditure is contained in Appendix C. The three main variables calculated are *tothhx1* (total household expenditure), *tothhx2* (total household expenditure without the imputation of the flow from durable assets), and *tothhx3* (total household expenditure without the flow from durables and imputed value of housing).

7.2 *Programs to calculate additional data files*

The additional data files have been generated using a set of programs in STATA. The series of programs and their layout is explained in Appendix E. These programs are supplied with the understanding that this is the only documentation that will be provided.

8. Notes to Potential Users

8.1 *Disclaimer*

The data have been collected using a specific sampling design and users should use the weights that take into account both the weighting and non-response. Using the current weights will produce statistics based on the estimated number of households. If the weights are multiplied by the number of individuals in the households they will reproduce the population in the sampling frame. The data can be regarded as representative only up to a certain level of disaggregation that contains a minimum number of cases.

Several cleaning procedures have been performed. The actual completed questionnaires were checked in the field by the supervisors. The data entry program controlled for typing errors and logical errors. Additional programs have been run by the World Bank. Still, there are probably outliers and inconsistencies in the data that have not been discovered or that have not been modified. In fact the principle utilized has always been to modify the data only if the new value can be trusted to be true. This is the case for typing errors and column shifts. Ultimately, it is up to the person analyzing the data to decide a methodological strategy to deal with missing values and outliers.

8.2 *Areas of difficulties and data problems*

One of difficulties encountered during the analysis of the data was represented by the weighting procedure. In particular, in one commune the derived weight was much bigger than the other weights. This meant that the households from that area would represent a large percentage of the sample and influence all the results. The procedure that is usually suggested to correct for such bias, and that was adopted here, is to replace the extreme weight with the second largest value (the weight for commune 22 was originally equal to 5531 and it was replaced with the second largest value equal to 1230). The immediate consequence of the adjustment is that the projected number of households becomes smaller than before. At this point there are two alternatives: one is to use the weights as they are and note that there is an undercount, the other is to expand all the weights in the rural areas to take into account the weight adjustment. The adjustment would be equal to the ratio of the projected rural households obtained using the unmodified weight to the number of projected households resulting from using the adjusted weights which is equal to 1.34.

The weight variables included in the data set do not include the extreme weight and were not adjusted. Any data user that would like to increase the weight of the households in rural areas can do so by multiplying the weights of the rural communes by 1.34.

Appendix A - How to Obtain the Data

The data are the property of the Minstry of Labour and Social Affair (formerly Ministry of Labout and Social Protection) of the Government of Albania. Those interested in using the data should contact:

Valentina Leskaj
Minister of Labor and Social Affair
Rruga e Kavajes
Tirana, Albania
tel 355-4-251351
fax 355-4-228340

Individuals requesting copies of the data should provide a brief description of the studies they plan to do and indicate that they will ask the LSMS Office to provide copies of the data and documentation.

After receiving permission from the Government, the LSMS Office at the World Bank will be able to distribute the data to prospective users by contacting:

LSMS Database Admnistrator
Development Research Group
The World Bank
1818 H Street, NW
Washington, DC 20433 USA
fax: (202) 522-1153
e-mail: lsms@worldbank.org
WWW: http://www.worldbank.org/lsms/lsmshome.html

Individuals requesting copies of the data from the LSMS Office should provide the following information:

- a copy of the permission from the Albanian Government to use the data;
- a brief description of the work they plan to do;
- an indication of the format in which they prefer to receive the data (ASCII, SAS Portable, STATA); and
- a check made out to the World Bank for the processing fee.

There is a nominal fee associated with the data sets. The World Bank provides them on 3½" diskettes in SAS Portable, STATA (version 2.1) or ASCII formats. The Development Research Group, Poverty and Human Resources requests copies of all reports and documents resulting from research that uses the data. The researcher should further note that once received, the data cannot be passed on to a third party for any reason. Other researchers must contact the MOLSP for permission to use the data. Any infringement on this policy will result in the denial of future access to World Bank LSMS data.

Appendix B - Selection of Households and Calculation of Weights

i) Selection in Rural areas

									Selection			Enumeration			Sample			Weights	Projected N Hhs	
bashki	Num	Region	Commune	Pop	Pop Rec.	% Rec	Num Hhs	Hh size	R Num	Start	Interval	Hh count	Change	Complete	Not Com	T Hhs	Hhs Rec. Nd.	Weight	Receive	Non Rec.
1	1	Berat	Bogove	3,189	790	24.8	742	4.30	0.250	7	25.6	804	1.08	26	6	32	15.38	302	1210	6654
2	14	Berat	Vendres	2,685	1,042	38.8	680	3.95	0.230	6	23.4	700	1.03	28	2	30	14.29	183	732	4390
3	20	Diber	Arras	6,194	4,273	69.0	1750	3.54	0.137	9	60.3	1596	0.91	23	6	29	47.83	135	1486	1621
4	24	Diber	Fushe M	8,035	3,222	40.1	1100	7.30	0.997	38	37.9	1156	1.05	25	5	30	52.00	107	1393	1286
5	27	Diber	Kastrio	8,816	4,751	53.9	2224	3.96	0.361	28	76.7	2435	1.09	30	2	32	26.67	131	1045	2874
6	33	Diber	Luzni e	6,000	3,209	53.5	1168	5.14	0.329	14	40.3	1102	0.94	28	1	29	35.71	99	986	1775
7	34	Diber	Maqella	11,881	7,557	63.6	3058	3.89	0.735	78	105.4	3301	1.08	27	2	29	40.74	114	1251	1819
8	38	Diber	Qender	9,074	4,381	48.3	2878	3.15	0.546	55	99.2	2789	0.97	29	1	30	13.79	178	711	4445
9	44	Diber	Zall Da	3,813	2,860	75.0	1145	3.33	0.123	5	39.5	1141	1.00	27	2	29	44.44	113	1350	1688
10	29	Diber	Klos	11,773	3,038	25.8	2794	4.21	0.330	32	96.3	2912	1.04	25	4	29	16.00	279	1116	5861
11	48	Diber	Mucukull	4,737	2,374	50.1	1472	3.22	0.136	7	50.8	1124	0.76	11	5	16	18.18	236	472	2124
12	37	Diber	Ostren	6,706	3,313	49.4	1519	4.41	0.082	5	52.4	1671	1.10	25	4	29	24.00	139	835	2644
13	41	Diber	Shupenz	7,375	3,817	51.7	2030	3.63	0.959	68	70.0	2116	1.04	22	7	29	22.73	183	917	3118
14	52	Durres	Cudhi	4,415	2,538	57.5	1270	3.48	0.324	15	43.8	1309	1.03	21	3	24	66.67	150	2095	1047
15	65	Elbasan	Paper	7,475	1,682	22.5	1610	4.64	0.893	50	55.5	2105	1.31	26	3	29	15.38	279	1117	6145
16	75	Elbasan	Gjinar	4,744	1,401	29.5	1150	4.13	0.181	8	39.7	943	0.82	21	8	29	14.29	297	890	5338
17	84	Elbasan	Kokovjat	4,647	1,050	22.6	1000	4.65	0.648	23	34.5	1062	1.06	15	8	23	0.00	382	0	5731
18	88	Elbasan	Lenie	2,958	1,429	48.3	678	4.36	0.644	16	23.4	658	0.97	28	0	28	25.00	124	869	2607
19	98	Elbasan	Prenjas	14,942	4,134	27.7	2345	6.37	0.833	68	80.9	3696	1.58	20	7	27	10.00	200	401	3606
20	100	Elbasan	Qukes	10,162	3,827	37.7	2225	4.57	0.782	60	76.7	2341	1.05	24	5	29	33.33	184	1470	2941
21	106	Elbasan	Sterblev	2,754	1,314	35.4	673	4.09	0.595	14	23.2	570	0.85	24	3	27	16.67	151	603	3015
22	130	Fier	Mbrosta	8,002	115	1.4	2107	3.80	0.185	14	72.7	2118	1.01	26	4	30	0.00	5531	0	143805
23	168	Gjirokas	Luftinje	6,341	452	19.3	1413	4.49	0.741	37	48.7	1386	0.98	21	9	30	4.76	1168	1168	23368
24	174	Korce	Cerrave	9,016	3,079	34.2	2557	3.53	0.725	64	88.2	2845	1.11	25	3	28	4.00	243	243	5840
25	198	Korce	Velcan	5,197	2,091	40.2	1498	3.47	0.795	42	51.7	1597	1.07	22	8	30	9.09	256	511	5112
26	190	Korce	Pirg	8,439	671	7.9	2063	4.09	0.307	22	71.1	2285	1.11	24	6	30	0.00	1005	0	24131
27	204	Kukes	Bicaj	8,608	5,073	58.9	2005	4.29	0.326	23	69.1	2036	1.02	19	11	30	68.42	163	2122	980
28	206	Kukes	Bushtric	11,967	9,372	78.3	2190	5.46	0.679	52	75.5	3000	1.37	21	9	30	90.48	87	1659	175
29	212	Kukes	Kolsh	2,069	911	44.0	480	4.31	0.265	5	16.6	409	0.85	21	9	30	4.76	197	197	3939
30	215	Kukes	Malzi	7,948	2,806	35.3	1727	4.60	0.181	11	59.6	2036	1.18	15	13	28	53.33	301	2405	2104
31	218	Kukes	Shishtav	7,192	4,174	58.0	1300	5.53	0.083	4	44.8	1638	1.26	20	10	30	50.00	122	1222	1222
32	221	Kukes	Terthore	4,647	2,422	52.1	960	4.84	0.052	2	33.1	907	0.94	20	10	30	50.00	156	1556	1556
33	224	Kukes	Ujemisht	4,916	3,329	67.7	1200	4.10	0.610	26	41.4	1145	0.95	14	16	30	100.00	202	2829	0
34	208	Kukes	Fajze	4,799	1,577	32.9	1185	4.05	0.233	10	40.9	1186	1.00	22	8	30	36.36	268	2145	3753
35	230	Lezhe	Mamurra	14,249	3,193	22.4	4539	3.14	0.005	1	156.5	5266	1.16	28	2	30	14.29	398	1594	9564
36	234	Lezhe	Sellte	5,090	1,900	37.3	1485	3.43	0.486	25	51.2	1324	0.89	14	16	30	21.43	438	1315	4820
37	239	Lezhe	Kacinar	4,142	2,141	51.7	1180	3.51	0.476	20	40.7	1143	0.97	18	12	30	38.89	240	1682	2644
38	254	Shkoder	Lac Van Dejeis	9,157	2,280	24.9	2239	4.09	0.022	2	77.2	2769	1.24	21	7	28	4.76	343	343	6851

39	256	Shkoder	Postrlb	10,014	5,860	58.5	3307	3.03	0.082	10	114.0	3174	0.96	18	3	21	44.44	172	1378	1723
40	258	Shkoder	Rrethin	13,005	4,639	35.7	3555	3.66	0.185	23	122.6	4294	1.21	25	5	30	12.00	241	722	5293
41	263	Shkoder	Pult	4,179	3,824	91.5	1040	4.02	0.886	32	35.9	1142	1.10	18	12	30	77.78	119	1660	474
42	247	Shkoder	Flerze	4,865	3,526	72.5	1399	3.48	0.960	47	48.2	1383	0.99	10	13	23	50.00	239	1194	1194
43	248	Shkoder	Fushe a	13,570	4,594	33.9	4525	3.00	0.761	119	156.0	4198	0.93	13	19	32	30.77	634	2537	5709
44	252	Shkoder	Kastrat	9,903	1,968	19.9	2522	3.93	0.475	42	87.0	2266	0.90	22	8	30	22.73	457	2286	7772
45	269	Shkoder	lballe	4,737	3,108	65.6	1444	3.28	0.276	14	49.8	1447	1.00	15	10	25	40.00	203	1216	1823
46	277	Tirane	Kashar	12,959	2,192	16.9	2970	4.36	0.121	13	102.4	4361	1.47	24	6	30	12.50	443	1329	9305
47	288	Tirane	Vaqar	6,311	300	4.7	1175	5.37	0.152	7	40.5	1678	1.43	25	5	30	8.00	1230	2459	28282
48	294	Tirane	Zall Bas	7,713	4,590	59.5	1906	4.05	0.514	34	65.7	2311	1.21	24	7	31	45.83	140	1544	1824
49	307	Vlore	Vranish	5,403	2,139	39.6	1316	4.11	0.445	21	45.4	1361	1.03	16	12	28	12.50	282	563	3944
50	92	Elbasan	Orenje	9,850	3,325	33.8	1891	5.21	0.925	61	65.2	2159	1.14	25	5	30	8.00	179	357	4107
TOTAL Selection				366,663	147,653	40.27	90689					98395	1.08	1,091	342	1,433	27.77		59,186	382,043
Total Rural Areas				2,004,082	379,362	18.93	496060													

		Tot	441,230
		% Receiv	13.41

Total Take	1450
# Comunes	50
Average Take	29
Total Hhs	90689
Average Interval	62.5
HHsize	4.06

P(ab) for the first commune is
50 * 790/379,362 * 29/742 * 26/32

ii) Selection in Urban areas

Bashki	Num	Prefect	District	Bashkie	Pop	# Recip	% Recip	Blocks	Estim Hhs	Actual Hhs	Int	Random	Start	Complete	Hhs %ND	weight	Projected Rec.	Non Rec
61	43	Lezhe	Laci	Laci	20,682	5,714	27.6	16	323	4799	5.4	0.49335	3	55	10.91	326	1,957	16,147
62	7	Elbasan	Elbasan	Elbasan	87,711	19,790	22.6	64	342	22262	5.7	0.90305	6	50	10.00	481	2,403	22,559
63	28	Korce	Pogradec	Pograde	21,242	5,045	23.8	16	331	5010	5.5	0.30149	2	49	6.12	433	1,298	18,765
64	13	Korce	Korce	Korce	65,451	10,386	15.9	64	255	18181	4.3	0.14965	1	52	11.54	719	4,315	31,856
65	38	Vlore	Vlore	Vlore	71,069	10,421	14.7	64	277	17292	4.6	0.30376	2	54	14.81	656	5,250	32,770
66	3	Korce	Devoll	Blilsht	5,424	670	12.4	1	1,354	1356	22.6	0.16973	4	53	7.55	816	3,264	33,255
67	35	Shkoder	Shkoder	Shhoder	82,097	22,747	27.7	64	320	20627	5.3	0.47432	3	52	23.08	373	4,471	18,135
68	17	Kukes	Kukes	Kukes	24,799	7,573	30.5	16	387	4940	6.4	0.41565	3	47	6.38	296	889	13,935
Total Selection					378,475	82,346	21.8		94,467	94,467				412	11.41		23,847	187,422
Total Urban area					926,563	170,893	18.4		231,269									
																	TOT	211,270
																	%Rec	11.29

Total Take	480
# Bashki	8
Average take	60
Total Hhs	
Average Interval	
Average Hh Size	4.01

APPENDIX C - Calculation of Welfare Measures - Total Consumption

The EXPAGG data file contains a measure of welfare based on the estimation of total consumption, calculated as the sum of all expenditures in cash and in kind that were reported. Calculating a measure of welfare based on consumption involves a series of data manipulations, handling of outliers and other assumptions to be made along the way. The methodology used and the assumptions made should be clear from the notes in the programs. Any researcher that does not agree with any of the assumptions made is encouraged to take a close look at the programs and to make his/her own estimations or to modify the existing programs included with the data. These programs and this appendix are provided with the understanding that no additional documentation will be provided. The organization of the programs used to calculate total aggregate expenditure is reported in table C1. The programs themselves are available with the data.

Table C1. Layout of Expenditure Programs

Program	Description of program	Input files	Output files
FDPRICE	Analyze food price data , decide on outlier correction, create data set with new price	corehh , sec10	fdprice
FDEXP	Create food expenditure and quantity data	corehh, sec10, fdprice	fdquant, fdexp
NFDEXP	Calculate non food expenditure	corehh, sec07b2 sec091, sec111 sec112, sec121 sec122	nfdexp
DURABLE	Calculate flow of services from durables	corehh, sec093	durable
RENTIMP	Hedonic rent regression and create imputed rent variable	corehh, sec091	rentimp
EXPAGG	Construction and analysis of aggregate expenditure	corehh fdexp nfdexp durable rentimp	expagg

The measure of total aggregate expenditure calculated includes four main components: food expenditure, non-food expenditure, housing expenditure and the estimation of the flow of consumption from durables. These components are discussed here in more detail.

a) The food expenditure variable includes the values and quantities of food items purchased, produced or received from other sources and were consumed during the previous month. They were evaluated at the last market price paid or known by the respondent. Extreme care was used to check the consistency of the price values. Outliers were identified, checked and replaced with medians relative to the same commodities (see program FDPRICE). Outliers were defined as larger than the (mean+3*sd and >3*mean) or less than the mean-3*sd) or (larger than the mean+4*sd and >4*mean) or (less than the mean-4*sd).

b) Non-food items include several expenses from different data files. There are: monthly expenditures (on such personal items as cigarettes, entertainment, transport cost and miscellaneous costs) from Section 11.1; yearly expenditures (such as personal items, clothing and other items) from Section 11.2; health expenditures from Sections 11.2 and 12 (the expenses from chronic and acute illness were used if they were larger that the annual amount reported in the yearly expenditure section); remittances from Section 7.2; and other housing expenses (utilities like electricity, water telephone, etc.).

c) Flow of expenditure estimates from durables. Households receive a consumption value from the household durables they own. An estimate of this value has been made using the information available in Section 9.3. The value itself has been calculated as the current value of the item divided by the expected life span of the item, which is equal to the average age of the items in the sample times two.

c) Imputed rental value. Most households in Albania own the dwelling in which they live (96 percent). In the questionnaire they were asked to report the value of their dwelling and the estimate rental value for a similar place. Estimates of rental values were received for approximately 90 percent of all households. A hedonic regression analysis was used to estimate the rental value of the remaining households. The predicted values have been used for all the households to smooth the estimates provided by the household themselves.

Three total aggregate measures have been calculated. The first (*tothhx1*) includes all the components described above. The second (*tothhx2*) does not include the estimate of consumption stream from household durables. The third (*tothhx3*) does not include the estimate from durables and the imputed value of rent.

Appendix D - Other Documents Available Upon Request

The following documents can be obtained from the World Bank, Poverty and Human Resources, Development Research Group (DECRG) at a cost of five cents per page for photocopying.

A. Questionnaires (free of charge)

1. Employment and Welfare Survey

B. Other Documentation

1. Basic Information Document (23 pages)
2. Codebook (46 pages)
3. Data Checking Program (32 pages)
4. Descriptions of Household and Individual Files (46 pages)
5. Descriptions of Constructed Files (6 pages)
6. Basic Statistics (39 pages)

Appendix E - Programs to Calculate Additional Data Files

Name of Progam	*Purpose*	*Infiles*	*Outfiles*
Data preparation			
CONVERT2		alban*.raw	sec*.dta
DUPLIC		sec*.dta	
MKHHID		sec*.dta	
DESCRIBE		sec*.dta	
Household and individual cross cutting characteristics			
ALBCORE	Create houshold and individual data sets with core characteristics	sec00, sec011 sec081, sec091 weights	corehh, coreind
NE_TEST	Test sampling weights	corehh, sec081	
Expenditure construction			
FDPRICE	Analyze food proce data , decide on outlier correction, create data set with new price	corehh , sec10	fdprice
FDEXP	Create food expenditure and quantity data	corehh, sec10, fdprice	fdquant, fdexp
FOODOUT	Analyze food quantity outliers	fdquant	
CALCON	Create calorie conversion dataset	fdquant	kcalfd
CALAN	Calculate per capita calorie dataset	fdquant, kcalfd	pccal
DURABLE	Calculate flow of services from durables	corehh, sec093	durable
RENTIMP	Hedonic rent regression and create imputed rent variable	corehh, sec091	rentimp
NFDEXP	Calculate non food expenditure	corehh, sec07b2 sec091, sec111 sec112, sec121 sec122	nfdexp
NFOODOUT	Analyze non food expenditure outliers	nfdexp, durable	
EXPAGG	Construction and analysis of aggregate expenditure	corehh fdexp nfdexp durable rentimp	expagg
Analysis of Expenditure			
EXPANN	Analyse Expenditure data		
PEXPAN	Analyse predicted expenditure data		
PEXPREG	Predicted expenditure regressions	pexpdath	preexp
PEXPPRED	Analyse changes in predicted expenditure between years	pexpdath	
PEXPFIT	Fit of predicted expenditure regressions	pexpdath	
Predicted expenditure regressions			
DEMOG	Construct demographic data	coreind, corehh sec012	demog
EDUC	Construct education data	coreind, corehh sec012, sec02	educ
NFASSETS	Construct non-farm business asset data	corehh, sec052a, sec052b	nfassets
LAND	Construct farm assets (land) data	corehh, sec062b, sec062c	land
MACHINES	Construct farm assets (machines) data	corehh, sec063	machines

LIVESTOCK	Construct farm assets (livestock) data	corehh, sec064	livestock
NEEXP	Construct social assistance and expemnditure data	corehh, sec081 expagg	neexp
LFORCE	Construct labor force data	adult, labhist	lforce
PEXPDAT	Aggreagate data for predicted expenditure regressions	demog, educ nfassets, land machines, livestck neexp, lforce	pexpodat
NE regressions			
ENTITLE	NE entit;ement construction	coreind, corehh sec031, pexpdat	entitle
NEREG	Construction of NE regression data. Note the F-test for identification are included in this program	(data from Ann Case, pexpdat, entitle, durable, rentimp	nereg
DIFF96R	Receipts of NE regressions first differences - rural 96	nereg	
REG95R	NE regression -- 1995 Rural	nereg	
REG95U	NE regression -- 1995 Urban	nereg	
REGHET95	NE regression -- 1996 Rural. Includes Hausman test and variations	nereg	
REGHET96	NE regression -- 1996 Urban	nereg	
TOB96R	Tobit estimates of NE receipt (including Hubert correction estimnates) -- 1996 rural	nereg	
Labor data			
ADULT	construct adult characteristics	coreind, corehh sec02, sec031 sec032, sec033 sec041, sec051 sec061	adult
LABMERGE	Check merging between roster and labor section	coreind, sec031 sec032, sec033	
LABHIST	Create labor history data	adult	labhist
Miscellaneous			
CHILDSUB	Extract information on household with children	sec011	chldsub
ROSTERAN	Analyze roster data	coreind	
SOCASSAN	Analyze social assistance data	corehh, sec031 sec081, sec082 sec121, sec122	
INCOME	copnstruct basic income data	corehh, sec041 sec051, sec062a sec71b, sec081 sec092b	income
INV*	Miscellaneous temporary investigation files		
CASESUB	Data from Anne Case	corehh, pexpdath	casesub