

Codebook for the “Life in Kyrgyzstan” Survey

Compiled by Susan Steiner with the help of Philipp Jäger

1 Background

The Life in Kyrgyzstan (LIK) survey has been conducted within the research project entitled “Economic Transformation, Household Behaviour and Well-Being in Central Asia: The Case of Kyrgyzstan” and funded by the Volkswagen Foundation. This project aims to (1) to collect nationally representative panel survey data in Kyrgyzstan, (2) to investigate well-being and household behavior in this country, and (3) to improve research capacity within the Central Asian region. It runs from January 2010 through December 2012 and is a collaborative endeavour of DIW Berlin, Humboldt University of Berlin, the Center for Social and Economic Research (CASE-Kyrgyzstan), and the American University of Central Asia (AUCA), the latter two being based in Bishkek, Kyrgyzstan.

The LIK is a multi-purpose, socio-economic survey covering a wide range of topics for economic and sociological research. Its principal objective is to provide data for the analysis of well-being. The data is broad enough to allow for different measures of well-being, such as, for example, income, child health, or life satisfaction. Two waves of the survey have been completed so far. Data was collected between October and December 2010 for the first wave and between September and November 2011 for the second wave.

All adult members of the households, not just one respondent, are interviewed and tracked in order to obtain a more complete picture of the household. This implies that if a member of an original sample household leaves the household (e.g. to form an own family), she is still part of the sample. If relevant, other members (e.g. spouses and children) of the new household are then part of the sample as well.

It is important to note that a household is defined in the LIK as all members of a domestic unit who normally live together, eat their meals together, and share their expenses. This is in line with Living Standard Measurement Study (LSMS) instructions in the region and is not trivial, as multiple family households are commonplace. In other words, households may consist of several families. They may also include individuals who are absent at the time of the survey due to study, work, or visits. This latter point is crucial as seasonal migrants may be listed as household members, even though they are abroad, if they would normally live, eat, and share the expenses with the rest of the household.

2 Sampling

The LIK collects data in all seven Kyrgyz oblasts (i.e. Batken, Chui, Djalal-Abad, Issyk-Kul, Naryn, Osh, and Talas) and the cities of Bishkek and Osh. It is representative at the national level as well as for urban and rural areas and for the south and the north of the country.

The original sample that was drawn for the first wave of data collection consists of 3,000 households and slightly more than 8,000 individuals in these households. The households were drawn through stratified two-stage random sampling. The strata are formed by Bishkek, Osh city, and the rural and urban areas of the seven oblasts, amounting to a total of 16 strata. At the first stage, a set of so-called population points (i.e. communities in rural areas, quarters in urban areas) were drawn in each stratum according to probabilities proportionate to population size. Table 1 provides an overview of the number of households in the total population and in the sample drawn for the first wave of the LIK. In each population point, a set of 25 households were drawn at the second stage. The National Statistical Committee (NSC) of the Kyrgyz Republic provided us with a household survey sample of 3,000 households based on the 2009 Population Census data. NSC also prepared reserve samples, ranging from 20 percent of the sample in rural areas to 100 percent in violence-affected areas in the southern part of the country. Whereas 73 percent of the households on the original sample list were found and interviewed; 27 percent of households had to be drawn from the reserve samples. Refusal to participate was quite high: Almost two thirds of the households from the original sample list that had to be replaced refused to be interviewed.

Table 1: Household population in Kyrgyzstan and first-wave LIK sample

	Number of households in total population, in '000 (in percent)			Number of households in first-wave sample (in percent)		
	All	Urban	Rural	All	Urban	Rural
Batken oblast	80.1 (6.99)	23.0 (2.01)	57.1 (4.98)	225 (7.50)	75 (2.50)	150 (5.00)
Chui oblast	206.3 (18.00)	48.5 (4.23)	157.8 (13.77)	550 (18.34)	125 (4.17)	425 (14.17)
Djalal-Abad	186.4 (16.27)	51.9 (4.53)	134.5 (11.74)	475 (15.84)	125 (4.17)	350 (11.67)
Issyk-Kul oblast	103.2 (9.01)	35.4 (3.09)	67.8 (5.92)	275 (9.16)	100 (3.33)	175 (5.83)
Naryn oblast	51.1 (4.46)	9.0 (0.79)	42.1 (3.67)	125 (4.16)	25 (0.83)	100 (3.33)
Osh oblast	187.9 (16.40)	15.3 (1.34)	172.6 (15.06)	475 (15.83)	25 (0.83)	450 (15.00)
Talas oblast	44.2 (3.86)	8.4 (0.73)	35.8 (3.13)	125 (4.16)	25 (0.83)	100 (3.33)
Bishkek	229.1 (19.99)	228.4 (19.93)	0.7 (0.06)	600 (20.00)	600 (20.00)	0
Osh city	57.5	53.5	4.0	150	150	0

	(5.02)	(4.67)	(0.35)	(5.00)	(5.00)	
Total	1.145.8	473.4	672.4	3000	1250	1750
	(100)	(41.32)	(58.68)	(100)	(41.66)	(58.33)

Source: The number of households in the population is from the 2009 Population Census.

3 Organization of Fieldwork

The survey is implemented by the data collection company Sotseconik based in Bishkek, Kyrgyzstan. This company has an excellent reputation in the region for conducting household surveys and has working experience with the World Bank and the Asian Development Bank. The interviews are held by around 120 recruited interviewers, who are supervised by 40 field supervisors. The field supervisors along with eight regional supervisors are responsible for logistical and administrative issues, as well as for counseling and data quality checks. They are also responsible for collecting information in the community questionnaire.

A pilot survey is run about one month before the main survey starts in each year. It is usually conducted in Bishkek (urban population) and Chui oblast (rural population), covering 30 households in total. Based on the pilot test results and consultations with the interviewers, the questionnaires are optimized to make them unequivocally clear. Training for all field staff, i.e. interviewers and supervisors, is then conducted. The training consists of explaining the goals of the project, in-class study of questionnaires, and in-class exercises. In the main survey, average workload per interviewer is approximately 25 households, but the actual time spent for interviews differs depending on the size and the situation of the household. On average, the time spent on one household questionnaire is approx. 45 minutes and on one individual questionnaire about 50 minutes. All interviewers and field supervisors are equipped with an interviewer manual that contains explanations of the field works process and the survey questions. A first data quality check is done by field supervisors when interviewers submit the filled questionnaires to them. The field supervisors re-interview about 5 percent of the households covered under their supervision. The second quality check is made at the data entry level. If there are major inconsistencies and missing information, the interviewers are asked to visit the same household again to correct answers.

4 Survey Instruments

The LIK survey consists of a household questionnaire (to be filled in by the most informed household member), an individual questionnaire (to be filled in by all adults of age 18 and above of the sampled households), and a community questionnaire (to be filled in by a representative of the local administration). Children (up to the age of 17 years) are not

respondents on their own. Information about them is collected in the household questionnaire. All questionnaires are first developed in English and are then translated into Kyrgyz and Russian.¹ The interviews are conducted as paper and pencil interviews in Kyrgyz or Russian, depending on the language preference of the interviewees. All interviewers are able to speak both languages.

Each of the questionnaires consists of several modules, all of which have been incorporated in the first two waves of the survey and will also be included in any future waves. Most of the modules consist, in turn, of several sections. Some changes were made to the questionnaires after the first wave. A few sections were dropped because they turned out to deliver too few useful observations (for example, Employees in Agricultural Markets, and Reverse Remittances) or were no longer relevant in 2011 (for example, Displacement). A number of sections that promised to provide information on interesting and relevant topics were newly included in the second wave (for example, Expenses on Customs and Traditions, Decision-Making, Time Use, Women's Background and Fertility). Within modules, improvements to the questions were also made but this was kept to a minimum in order to secure comparability of the data over time. In the following, the structure of the questionnaires is presented. The italic sections were included either in 2010 or in 2011 but not in both years.

Household Questionnaire

1. Household Roster
 - 1.1. Household Composition
 - 1.2. Child Education
 - 1.3. Child Health
2. Housing and Assets
 - 2.1. Housing
 - 2.2. Assets
 - 2.3. Land
3. Agricultural Markets
 - 3.1. Trade Activity
 - 3.2. *Employees*
 - 3.3. Quality Requirements
4. Consumption and Expenditure

¹ The questionnaires in all three languages can be accessed at <http://www.diw.de/kyrgyzstan>.

- 4.1. Food Items
- 4.2. Non-Food Items
- 4.3. *Expenses on Customs and Traditions*
- 5. Income Sources
- 6. Migration
 - 6.1. Current Labour Migration
 - 6.2. Remittances
 - 6.3. *ReverseRemittances*
- 7. Shocks

Individual Questionnaire

- 1. Subjective Well-Being
- 2. Education and Health
 - 2.1. Education
 - 2.2. Health
- 3. Labour Market
 - 3.1. Current Employment Status
 - 3.2. Work during the last 7 Days
 - 3.3. Current Unemployment or Inactivity
 - 3.4. Work during the last 12 Months
 - 3.5. Labour Market CV
- 4. Movements
- 5. Family and Household
 - 5.1. Family
 - 5.2. *Decision Making*
 - 5.3. *Time Use*
 - 5.4. *Women's Background and Fertility*
- 6. Worries
- 7. Security and Violence
 - 7.1. Perception of Security
 - 7.2. Violence
 - 7.3. *Displacement*
- 8. Social Life
 - 8.1. Membership in Groups
 - 8.2. Informal Networks
 - 8.3. Trust and Information

Community Questionnaire

1. Community Information
2. Prices for Food Products

5 Panel Design

One of the most crucial features of a longitudinal survey such as the LIK is to cope with problems of tracking and attrition. In principle, all persons who took part in the first wave of the survey in 2010 are to be surveyed in the following waves. Hence, individuals and not households are being tracked. In each survey year, all individuals aged 18 and older, who have been part of an LIK household in previous years, and their respective households are to be interviewed. In case of residential mobility, the respective person is to be followed within the survey territory (i.e. the Kyrgyz Republic). Third persons moving into an existing LIK household are to be surveyed and tracked, even in case of subsequently leaving that household. Since all adult household members are to be re-interviewed individually in the LIK, the next generation is automatically taken into account: All children of LIK households become part of the sample at the time of turning 18 years.

With regard to the sample sizes, 8,160 individuals were interviewed in the first wave and 8,066 individuals in the second wave. Of the individuals included in the first wave, 7,364 (or 90.2 percent) could be re-interviewed in the second wave. With regard to households, 3,000 households were interviewed in the first wave and 2,863 in the second wave. Of the households included in the first wave, 2,856 (or 95.2 percent) were re-interviewed in the second wave.

6 Non-Response to Modules

In the first and second waves, non-response to specific modules was in a reasonable range. At the household level, non-response was at a level of less than ten households for each module except for the income module. Twenty nine (out of 3,000) households did not provide any information to this module in the first wave and 14 (out of 2,863) in the second wave. Given the sensitivity of income, we consider these numbers of non-response to be acceptable. At the individual level, non-response was low for the first four modules (subjective well-being, education and health, labour market, and movements). However, it has been between 30 and 75 individuals (out of 8,160) for the last four modules in the first wave and ranging from 20 up to around 400 individuals (out of 8,066) in the second wave.

7 Data cleaning

The data cleaning process is mainly the responsibility of DIW Berlin. It is generally focused on data labeling and management of data entry mistakes. Whenever possible, data inconsistencies, for example out-of-scale answers and duplicate household or personal identifiers, are solved in consultation with the data collection company. Some inconsistencies, which are not easily solved without strong assumptions, may remain in the data files. For example, a respondent reported to be male in the household roster may appear in a female-only module. Even though it appears likely that this person is actually not male but female, we would not re-define the gender in the household roster because it is equally possible that the answers provided in the female-only module are given for a different person in the household (and hence, the personal ID code is wrong). To keep ungrounded data manipulation to a minimum, the decision about further adjustment is left to the researcher.

8 Using the Data

File structure

Each of the three questionnaires that are part of the LIK survey (i.e. community, household, and individual) is divided into distinct modules. The data files are generally stored per module in STATA compatible dta-files. The abbreviations “id” or “hh” indicate that a data file refers to the individual or household questionnaire. The subsequent numbers and letters link the data to specific modules and sections (sub-modules). The community data is saved under the names “a0” in 2011 and “cm” in 2010. The tables in the Annex visualize the structure and names of the data files.

Merging

All household and individual data files contain the variable “hhid” which uniquely identifies the households in the sample. In addition, all individual data files as well as the first three household data files include an individual identifier “pid”. Combining the household and personal codes, “hhid” and “pid”, allows for uniquely identifying every individual in the sample. Hence, merging household and/or individual data files within a given wave may be done using these variables. In some modules, e.g. hh2b, hh2c, or id3d, the identifier variables do not uniquely define a unit, due to multiple answers per household and/or individual. In these cases, a transformation of the data (for example, reshaping) may be necessary before merging files. The community information can be combined with the remaining dataset using

the “cluster” variable. Since this variable is only available in the control card, a merge between the control card and the module of interest is indispensable.

Merging across waves is slightly more challenging. The “hhid” and “pid” variables do not necessarily identify the same individuals across waves. When individuals move and/or form new households, these new households are given a new hhid. When the household composition changes within a household, for example when the household head leaves the household, the remaining members are newly enumerated and receive new pid numbers. Therefore, the data file “cctogether”, based on information from the control card, documents all changes in the household and personal identifiers between the last and the current year. (Unfortunately, the information in the control card is not completely consistent. To check for attrition and newly surveyed units, you should rely on the identifiers rather than questions q3 and q5 in the second wave’s control card.) When an individual is surveyed for the first time, his/her “pid” in the previous wave is “0”. An individual that is not present in the current wave any more has a “pid” equal to “0” in the current year.

To facilitate the merging procedure some sample merging files can be found in the Annex of this Codebook.

ANNEX

LIK data files, waves 1 and 2

Household Questionnaire

Modules and Sections	Dataset filename (2011)	Dataset filename (2010)
1. Household Roster		
Household Composition	hh1a	hh1a
Child Education	hh1b	hh1b
Child Health	hh1c	hh1c
2. Housing and Assets		
Housing	hh2a	hh2a
Assets	hh2b	hh2b
Land	hh2c	hh2c
3. Agricultural Markets	hh3	hh3
4. Consumption and Expenditure		
Food Items	hh4a	hh4a
Non-Food Items	hh4b	hh4b
Expenses on Customs and Traditions	hh4c	---
5. Income Sources	hh5	hh5
6. Migration		
Current Labour Migration	hh6, hh6a	hh6a1, hh6a2
Remittances	hh6b	hh6
Reverse Remittances	---	hh6
7. Shocks	hh7	hh7

Individual Questionnaire

Module	Dataset filename (2011)	Dataset filename (2010)
1. Subjective Well-Being	id1	id1
2. Education and Health	id2	id2
3. Labour Market		
Current Employment Status, ...	id3	id3
Work during the last 12 Months	id3d	id3d, id3d_347
Labour Market CV	id3e	id3e
4. Movements	id4	id4
5. Family and Household		
Family	id5a	id5
Decision Making	id5b	----
Time Use	id5c	----
Women's Background and Fertility	id5d, id5dch	----
6. Worries	id6	id6
7. Security and Violence	id7	id7
8. Social Life	id8	id8

Community Questionnaire

Module	Dataset filename (2011)	Dataset filename (2010)
1. Community Information	ao_01	cm1
2. Prices for Food Products	ao_02	cm2

STATA do-files for merging

```
*****
***** Sample Do-File for the LIK-Survey *****
***** -Merging across modules within a given year- *****
*****
```

```
clear
```

```
***** Merging an ID-file to another ID-File *****
```

```
use "your directory/id1"
```

```
merge 1:1 hhid pid using "your directory/id3"
```

```
***** Merging a ID-file to a HH-File *****
```

```
use "your directory/id1"
```

```
merge m:1 hhid using "your directory/hh3"
```

```
***** Merging a HH-file to a Community file *****
```

```
use "your directory/cctogether"
```

```
rename _merge merge_cc
```

```
drop if pid==0 // Deleting all new individuals in 2011 //
```

```
merge 1:1 hhid pid using "your directory/hh1a"
```

```
rename _merge merge_hh_cc
```

```
merge m:1 cluster using "your directory/ao_01"
```

```
*****
***** Sample Do-File for the LIK-Survey *****
***** - Merging modules across time - *****
*****
```

```
clear
```

```
***** Merging an Individual-file to another in the previous year *****
```

```
use "your directory_2011\cctogether" // To merge over time you will always need the control card
that links household and individuals over time //
```

```
* Drop observations that hamper the merging process*
```

```
drop if pid==0 // To avoid duplicates values (pid_2010==0 for all individuals
that did not participate in 2011 ) //
```

```
merge 1:1 hhid pid using "your directory_2011\id1"
```

Rename the variable to avoid losing them across time

```
foreach var of varlist i100-i103 {
```

```
    rename `var' `var'_2011
```

```
}
```

```
rename ih2010 hhid_2010
```

```
rename idres10 pid_2010
```

* Drop observations that hamper the merging process*

```
drop if pid_2010==0                                // To avoid duplicates values (pid_2010==0 for all individuals  
that did not participate in 2010 ) //
```

```
rename _merge merge_id_cc
```

```
drop if merge_id_cc !=3                            // Only individuals who answered the id1 in 2011 //
```

```
merge 1:1 hhid pid using "your directory_2010\id1"
```