

Data Documentation

Deutsches Institut für Wirtschaftsforschung

2011

The Kyrgyz Integrated Household Survey (KIHS) A Primer

Damir Esenaliev, Antje Kroeger and Susan Steiner

IMPRESSUM

© DIW Berlin, 2011

DIW Berlin

Deutsches Institut für Wirtschaftsforschung

Mohrenstr. 58

10117 Berlin

Tel. +49 (30) 897 89-0

Fax +49 (30) 897 89-200

www.diw.de

ISSN 1861-1532

All rights reserved.

Reproduction and distribution

in any form, also in parts,

requires the express written

permission of DIW Berlin.

Data Documentation 62

Damir Esenaliev

Antje Kroeger

Susan Steiner

German Institute for Economic Research (DIW Berlin)

The Kyrgyz Integrated Household Survey (KIHS): A Primer

Berlin, Dezember 2011

1. Motivation for this primer

In the course of the research project “Economic Transformation, Household Behaviour and Well-Being in Central Asia: The Case of Kyrgyzstan” conducted at DIW Berlin, we analyse micro-level survey data from several countries in Central Asia, mostly from Kyrgyzstan.¹ One dataset which we have used extensively is the Kyrgyz Integrated Household Survey (KIHS) for the years 2003-2009. In this primer, we summarise our experience in working with this dataset. We thereby aim to facilitate other researchers’ understanding and use of the dataset, which is all the more important as there is no official documentation for data users.

2. Micro-level data sources in Kyrgyzstan

A large number of micro-level socio-economic surveys have been carried out in Kyrgyzstan since the country’s independence from the Soviet Union in 1991 (Table 1).

Table 1: Major household surveys in Kyrgyzstan

Dataset	Period covered	Collected by	Sample size (number of households)	Data publicly available?
Multipurpose Poverty Survey	1993	World Bank and NSC	1,937	Yes
Poverty Monitoring Survey	1996-1999	World Bank and NSC	1,951 - 2,979	Yes
Household Budget Survey	1996-2003	NSC	1,200 - 3,000	No
Demographic and Health Survey	1997	USAID and NSC	3,672	Yes
Household Energy Survey	1999	London Economics and NSC	3,006	No
Kyrgyz Integrated Household Survey	2003-2010	NSC	5,012	No, for sale only
Survey of Conflict Prevention and Cooperation	2004	World Bank	1,500	Yes
Multiple Indicator Cluster Survey	2006	UNICEF and NSC	5,200	Yes
Life in Transition Survey	2006, 2011	EBRD	1,000	Yes
Life in Kyrgyzstan Panel Survey	2010-2011	DIW Berlin	3,000	Yes

Several standard international and regional surveys, such as the Demographic and Health Survey (DHS), the Multiple Indicator Cluster Survey (MICS), and the Life in Transition survey, have been conducted. Predecessors of the KIHS are the Living

¹ Since January of 2010, the Department for Development and Security of the German Institute for Economic Research (DIW Berlin) has been conducting the research project “Economic Transformation, Household Behaviour and Well-Being in Central Asia: The Case of Kyrgyzstan” funded by the Volkswagen Foundation. The project is a joint endeavour of DIW Berlin, the Humboldt University of Berlin, the Centre for Social and Economic Research (CASE Kyrgyzstan) and the American University of Central Asia in Bishkek with DIW Berlin taking the coordinating role. One of the key components of the project is the collection of the panel survey data from 2010 to 2012. The survey is called Life in Kyrgyzstan (LIK) (<http://www.diw.de/kyrgyzstan>).

Standard Measurement Study (LSMS) type surveys conducted in the 1990s, i.e. the Multipurpose Poverty Survey in 1993 and the Poverty Monitoring Survey in 1996-1999, as well as the Household Budget Survey conducted annually between 1996 and 2003. These surveys had the objective to monitor the living standard of the population and to estimate poverty. Thus, the National Statistical Committee (NSC) has gained extensive expertise in survey data collection. Beside NSC, there are several private (consulting) companies and non-governmental organisations that carry out micro-level data collection, for example M-Vector, SIAR Bishkek, and Sotseconik.

3. Structure of the KIHS

The KIHS was introduced in 2003 by NSC with financial and technical support from the UK Department for International Development (DFID). Oxford Policy Management provided technical assistance to the NSC in the survey design and the first year data collection.

The KIHS has been conducted quarterly since its inception in 2003. It covers close to 5,000 households at each point in time and is thus the household survey with the largest sample size in the Kyrgyz Republic. The KIHS is a rotating panel with a maximum of one quarter of the sample being replaced each year (see below). The sample of the KIHS is drawn using stratified two-stage random sampling, based on the results of the 1999 population census. The country is divided into 15 strata, representing urban and rural dimensions of the seven oblasts and the city of Bishkek. Each cross-section is designed to be representative at the national, rural/urban, and oblast levels (for concerns about this, see below). It is unclear whether panel sub-samples are representative at any of these levels as well.

In line with the Household Budget Survey, the KIHS is designed to accurately measure consumption-based poverty in the country and to analyse the socio-economic dimensions of people's living standards. It is therefore very comprehensive in data on consumption and expenditure.

Different from most of the earlier surveys (for example, the Household Budget Survey), the KIHS also includes a labour force survey, which makes it an "integrated" survey, in order to allow for analyses of the link between labour market outcomes and poverty. The KIHS is today the major data source for the production of national statistics in Kyrgyzstan on living standards, labour market indicators, and food security.

Given the objective of the KIHS to provide data for measuring consumption based poverty, the survey collects very detailed information on households' consumption and expenditure. Every household fills in a diary on food consumption, covering more than 150 food items, for two weeks each quarter of the year. In addition, information on expenditure for non-food items and services is collected during the quarterly interviews. This approach allows for a very comprehensive picture of household consumption and expenditure, much more comprehensive than in most other household surveys.

The KIHS broadly consists of the following parts:

- Household composition (age, gender, marital status, ethnicity of household members),
- Education,
- Internal migration,
- Health status and access to medical services and medicines, and fertility,
- Labor force (employment status, sector of employment, etc.),
- Expenditures for food and non-food products, services, taxes and contributions, purchase of durables, and own production of goods,
- Income sources (such as wages, pensions, social payments, self-employment profit, remittances, etc.),
- Housing conditions and ownership of durables.

On the one hand, the KIHS includes a number of questionnaire modules and questions which are not standard practice. For example, a module on women's fertility is included and children are physically measured in the health module. On the other hand, given that labour migration and remittances play a central role for Kyrgyz households, there is only little coverage of these important topics in the KIHS.

4. Panel structure of the KIHS

The KIHS has a unique panel structure. A maximum of one quarter of households are replaced in each year, and in many years the replacement rate is substantially lower (Table 2). As a consequence, 31 percent of the households that had been part of the 2003 sample are still part of the 2009 sample. Related to this, it is relatively easy to build a household panel as the household ID codes uniquely identify the households.²

This provides the potential for panel data estimation techniques, which are crucial for correctly investigating welfare dynamics. There are not many panel surveys, and very few for more than two time periods, in the whole of Central Asia, which makes the KIHS an innovative and important dataset.

However, there are two serious concerns for the representativeness of the cross-sectional samples as well as the panel sub-samples. First, households are not tracked over time if they move within Kyrgyzstan. Such households are simply dropped from the sample and replaced in the following year. Given the high rate of internal migration in the Kyrgyz Republic, this potentially introduces a selection bias.

² In the light of our experience with another household survey from the Central Asian region, where household ID codes were re-assigned to newly sampled households, it is worth mentioning this decisive aspect, even though it might seem an obvious precondition for panel surveys.

Table 2: Panel structure of KIHS, 2003-2009

Year	2003	2004	2005	2006	2007	2008	2009
Number of households	4760	4626	4771	4863	4803	4995	4984
Panel of 2003-09		3961	2648	2426	1773	1650	1498
<i>in % from 2003</i>		83	56	51	37	35	31
Panel of 2004-09			3071	2809	2089	1944	1772
<i>in % from 2004</i>			66	61	45	42	38
Panel of 2005-09				4341	3384	3090	2820
<i>in % from 2005</i>				91	71	65	59
Panel of 2006-09					3817	3473	3153
<i>in % from 2006</i>					78	71	65
Panel of 2007-09						4353	3916
<i>in % from 2007</i>						91	82
Panel of 2008-09							4444
<i>in % from 2008</i>							89

Source: Authors' calculation from KIHS data for 2003-2009.

Second, it is unclear how many (and which) households are dropped from the sample in addition to those who moved or dissolved. It appears that a number of households are randomly selected to be dropped from the sample each year. Yet, given the differing replacement rate in different years (see Table 2), there seems to be no systematic process, which could consist, for example, in achieving the 25 percent replacement rate. It is also unclear how exactly the households that were dropped are replaced.

While household ID codes are unique and consistent over the years, the individual ID codes are not unique to individuals. Individuals are simply numbered serially in each household, which may give the same individual ID code to different individuals if there are changes in the household composition over time. In order to build a panel dataset at the individual level, however, it is necessary to identify the same individuals over time. In the KIHS, this is now only possible through merging datasets of different time periods based on household ID codes and time-invariant individual characteristics, such as gender and birth year. Yet, such an approach leaves substantial room for mistakes (for example, in the case of twins) and could easily be avoided by assigning unique individual ID codes to individuals at the time of data collection.

There are a number of data inconsistencies across the waves which may have happened during the data entry process. For example, there are different birth dates and ethnicities given in different years for an obviously identical person. Thirteen percent of all individuals who were interviewed in the three years between 2003 and 2005 provided varying answers for their ethnicity.

In several modules of the questionnaire, questions were changed or re-phrased from one year to another. Although this is often inevitable due to realities on the ground, it makes panel data analysis difficult.

In order to make full use of the panel structure, it would be desirable that the NSC started to look at the data from a panel perspective. This starts with naming the same variables in the same way, asking the same questions over time (which has been done since 2004 to a very large extent) and identifying individuals in households uniquely over time. Merging datasets of different years at NSC helps detect inconsistencies that could be easily corrected. Besides, any future documentation would gain in utility from including information on how to merge data sections over several years and how to create an individual panel.

5. Data quality

About one percent of households in the 2003 sample do not have a unique household head. Given that 2003 was the first year of the KIHS, it could be that interviewers considered both spouses as heads of household. The question remains how to decide who the real household head is, if both spouses are indicated as heads.

There are a lot more households in the dataset than there are in the so called *basic* data file of each year; hence, basic characteristics, such as oblast, urban or rural location, etc., which are included in this file, are missing for many households. For example, there are only 4,771 households in the *basic* data file but 4,966 households in the *fl* data file in 2005. Even though it would be easy to delete those observations without basic information from the dataset, this gives rise to concerns about representativeness of the remaining sample.

The reported total amount of monetary values is in some cases not equal to the sum of the different components of these values, and it is unclear where the deviation comes from. For example, the amount of total expenditures on education is not equal to the sum of the components, at least for the years 2005-2008. It would hence be desirable if the data collection and data cleaning processes included more double checking (by interviewers themselves, supervisors, and data entry staff) on reported answers of respondents, particularly for variables with financial terms.

6. Sampling weights

The calculation of sampling weights is not very clear. We observed that the weights increase on average over time, and more so for some groups of the population than for others. For example, we investigated the weights of different ethnic groups and found that the average weight for Uzbeks increased by a factor of 4.9 between 2003 and 2005, while the average weight for the other ethnic groups increased by about 2.7. According to information from the NSC, the calculation of weights takes into consideration the sampling probability of primary sampling units in each stratum and the sampling probability of households in these units. These weights are then modified in order to give a realistic picture of different age groups (children, working age adults, elderly) in the society, but they reportedly do not control for ethnicity. A disproportionate increase

of weights for Uzbek households would then only make sense, if the age structure of Uzbeks changed dramatically over time, if the number of Uzbeks decreased in the sample, or if Uzbek households became smaller and smaller. None of these options seems likely.

7. Research based on KIHS

Being the longest and largest panel household survey in the Kyrgyz Republic, the KIHS is an important data source for the analysis of socio-economic developments in the country. In addition to several policy reports using KIHS data (for example, World Bank 2003, 2005, 2007a, 2007b, 2007c), the following academic studies have analysed KIHS data.

Falkingham, Akkazieva and Baschieri (2010) examine the out-of-pocket health expenses in Kyrgyzstan after the health reform of 2002. This reform was implemented in order to increase the transparency of financial contributions by patients. Overall, the authors find that access to health care within the health sector in the Kyrgyz Republic has improved. However, the burden of outpatient health care payments amongst the poor remains significant. As a result of the introduction of co-payments for hospital care, fewer inpatients report making payments to medical personnel, but when they are made, payments are high, especially to surgeons and anesthetists. The overall out-of-pocket costs of inpatient care have fallen and equity has improved. They conclude that Kyrgyzstan provides a model that could be replicated throughout the region.

Ukueva and Becker (2010) study the impact of internal and international remittances on the purchase of durable goods. Using data for the years 2005-2007, they find that both international remittances and domestic transfers flow to more vulnerable families, such as female-headed households, or households with a higher proportion of elderly members. They also note that domestic transfers are generally more important in Northern regions, whereas remittances are more important in the South. Unlike internal transfers, international remittances in Kyrgyzstan flow to households with a higher income level. This suggests that international migration might be costly and low-income families cannot afford to send family members abroad. They come to the conclusion that remittances in Kyrgyzstan are used to support investments in durable goods.

The analysis by Kroeger and Anderson (2011) examines the impact of international remittances and domestic transfers on children's capabilities between 2005 and 2008. They find that remittances and domestic transfers have generally not promoted investments in the human capital of children. Specifically, preschool enrollments are higher in the urban north but secondary school enrollments are lower in other regions in remittance receiving households; expenditures are also negatively affected in the south and the mountain areas. These negative enrollment results are larger for girls than for boys. They also find evidence of stunting and wasting among young children and worse health habits among boys in remittance or transfer receiving households. In the long run, Kyrgyzstan needs human capital development for growth; this study suggests that

remittances are not providing the boost needed in human capital to promote development in the future.

Esenaliev and Steiner (2010) use data from three cross-sections of the KIHS (2003, 2005 and 2008) in order to measure horizontal inequality between ethnic Uzbeks and Kyrgyz in Kyrgyzstan. The motivation for doing so is the ethnically motivated violence in June of 2010, which was often explained in the media by substantial differences in the welfare of these two ethnic groups. The authors use three alternative measures of welfare, namely household consumption, a standard asset index, and a variable measuring the value of assets that are most visible (i.e. house and car). They find some evidence for higher welfare of Uzbek households compared with their Kyrgyz counterparts in the Southern regions, mainly in rural areas. The prevalence of such horizontal inequality has strong policy implications, as it implies the potential danger of future violence between the two groups.

8. Conclusion

To sum up, the KIHS is an important dataset that contains comprehensive information on the sample households, especially in terms of consumption and expenditure. Being a rotating panel, it provides the potential for panel data analysis at the household level. However, this potential has not been much used, neither by the NSC nor by researchers. In order to provide valuable input into the policy-making process in Kyrgyzstan, it would be desirable that this changes in the near future.

Bibliography

Esenaliev, Damir and Steiner, Susan (2010): Are Uzbeks better off? Economic welfare and ethnicity in Kyrgyzstan, Paper presented at Conference of the Verein fuer Socialpolitik Research Committee on Development Economics, Berlin, June 2010.

Falkingham, Jane, Akkazieva, Baktygul and Baschieri, Angela (2010): Trends in out-of-pocket payments for health care in Kyrgyzstan, 2001-2007. *Health Policy and Planning*, 25(5), 427-436.

Kroeger, Antje and Anderson, Kathryn (2011): Remittances and Children's Capabilities: New Evidence from Kyrgyzstan, 2005-2008. Case Network Studies and Analyses No 430/2011. Warsaw: Center for Social and Economic Research.

Ukueva, Nurgul and Becker, Charles (2010): Private transfers in Kyrgyzstan's post transition environment: results from a new household panel dataset. In: *Migration, Remittances, and Growth*. Dissertation by Nurgul Ukueva, p.59-83. Department of Economics, Duke University.

World Bank (2003): Kyrgyz Republic. Enhancing Pro-Poor Growth, Washington DC: World Bank.

World Bank (2005): Growth, Poverty, and Inequality: Eastern Europe and the Former Soviet Union, Washington DC: World Bank.

World Bank (2007a): Kyrgyz Republic Poverty Assessment: Growth, Employment, and Poverty, Washington DC: World Bank.

World Bank (2007b): Kyrgyz Republic Poverty Assessment: Labour Market Dimensions of Poverty, Washington DC: World Bank.

World Bank (2007c): Kyrgyz Republic Poverty Assessment: The Main Conclusions of the Report, Washington DC: World Bank.