Final Draft: June 17, 1996

THE 1993 KYRGYZSTAN MULTIPURPOSE POVERTY SURVEY: DOCUMENTATION

Poverty and Human Resources Division
Policy and Research Department
World Bank

Table of Contents

1. Background 1
2. Data Collected 2
   2.1. Household Questionnaire 2
   2.2. Adult Questionnaire 5
   2.3. Child Questionnaire 8
   2.4. Survey of Availability and Prices of Food Products and Fuel 9
   2.5. Survey of Community and Social Infrastructure 11
3. Sample 12
   3.1. The formation of strata 13
   3.2. The selection of primary sampling units 16
   3.3. The selection of secondary sampling units 16
   3.4. The selection of last stage sampling units (households) 18
   3.5. Price survey sample procedures 18
   3.6. Survey as implemented 19
4. Field Work 19
   4.1. Design, translation, piloting and printing of the questionnaires 19
   4.2. Organization of field work 20
   4.3. Interviewer training 21
   4.4. Field work 22
   4.5. Editing, coding, data entry and cleaning 22
5. Using the Data 22
   5.1. Data set and variable names 23
   5.2. Using the nutrition data 23
   5.3. Data quality 24
   5.4. Specific data issues 25
   5.5. Weighting factors 26
   5.6. Linking components of the data 26
6. Constructed Data Sets 27
   6.1. Household demographic and relationship variables 28
   6.2. Adult characteristics 29
   6.3. Household characteristics 32
   6.4. Household expenditure variables 34
   6.5. Household income variables 35
   6.6. Poverty line and poverty indicators 38

Appendices

Appendix A. How to Obtain the KMPS Data 40
Appendix B. List of KMPS Data Sets Available 41
Appendix C. List of Related Documents 42
Appendix D. List of Papers and Reports 43

Appendix E. The Poverty Line....44
Appendix F. Codes Not in the Questionnaires....47
Appendix G. Food Price Data....49

List of Tables

Table 1: Household Questionnaire.... 2
Table 2: Adult Questionnaire.... 5
Table 3: Child Questionnaire.... 9
Table 4: Stages of the sampling process....13
Table 5: Distribution of households in self-representing strata....14
Table 6: Distribution of households in non self-representing strata....15
Table 7: Constructed variables: household demographic and relationship variables....28
Table 8: Constructed variables: adult characteristics....30
Table 9: Constructed variables: household characteristics....33
Table 10:Constructed variables: household expenditure....34
Table 11: Constructed variables: household income....36
Table 12: Constructed variables: poverty line and poverty indicators....39
Table 13: Annual food basket and subsistence income level, October 1993....46
Table 14: High-cost food basket and subsistence income level, by oblast, October 1993....46
Table 15: Low-cost food basket and subsistence income level, by oblast, October 1993....46
Table 16: Kyrgyz Republic-wide market food prices, October 1993 (soms per kg)....51
Table 17: Prices for food items by oblast, October 1993....52

This report was prepared by Robert Ackland, consultant to the PRDPH Division. Comments were provided by Margaret Grosh, Barry Popkin, Kinnon Scott and Michael Swafford.

1. Background

In the 1980s the World Bank designed a survey mechanism to measure the living standards of the populations in developing countries. The data collected using a Living Standards Measurement Survey (LSMS) can be used to assess development policy, measure and identify poverty, and evaluate the effectiveness of various government interventions in the area of poverty alleviation, social services, labor and community infrastructure. The first of these LSMS were carried out in the Ivory Coast and Peru in 1985 and, in the past eight years, LSMS have been carried out in a variety of countries.[1]

The Kyrgyzstan Multipurpose Poverty Survey (KMPS) was designed to be a nationally representative survey capable of measuring the standard of living in the Kyrgyz Republic[2] during the second half of 1993. The survey was conducted in October and November 1993, and the sample contains about 2,000 households and 10,000 members of those households. In addition, there is information on the availability and prices of food products and fuel, and also data on community and social infrastructure. While the KMPS is based on the LSMS framework, it has some features which distinguish it from the standard LSMS; in particular it collects extensive nutrition data.

The tradition of survey research in countries of the Former Soviet Union (FSU) is not particularly strong. In the Kyrgyz Republic, the GOSKOMSTAT family budget surveys were not representative of the population in general, and the poor in particular. These surveys tended to focus on persons who work in enterprises and, to a lesser extent, pensioners. The KMPS represents a significant increase in the data available, and is a more suitable tool for monitoring the social and economic changes currently occurring in the Kyrgyz Republic.

The 1993 KMPS was carried out under the direction of researchers from the University of North Carolina at Chapel Hill, Paragon Research International, Inc., and the Institute of Sociology of the Russian Academy of Sciences.[3] The government of the Kyrgyz Republic has recently established an open access policy in regards to the data collected in the KMPS (for details, see appendix A). The potential uses of this data set are quite broad given the multi-topic nature of the data and the fact that it was carried out at the national level. The purpose of this paper is to provide detailed documentation of the KMPS in order to: a) simplify its use for potential users thereby lowering start-up costs to analysts; and b) ensure that the procedures used in the design, implementation and initial analysis of the survey are chronicled accurately. Such documentation will serve both to facilitate use of the data set and to prevent misuse of the data due to misunderstandings of the sample and/or field work procedures.

---

[1] For example: Peru, Ghana, Ivory Coast, Jamaica, Vietnam and Bolivia.

[2] After independence, the name of the country was changed from Kyrgyzia to Kyrgyzstan. Since then, the name has been changed again to the Kyrgyz Republic.

[3] Hereafter referred to as the 'survey team'. Four future rounds of the KMPS are planned, with the first round commencing in 1995.

The paper is organized as follows. Section 2 contains an outline of the types of data collected in the KMPS. Section 3 provides information on the sample design. A description of the field work is found in Section 4 and in Section 5 there is specific information about using the data. Information on constructed variables is provided in Section 6.

2. Data Collected

The KMPS consists of five components: a household questionnaire; an adult questionnaire; a child questionnaire; a food price and availability survey; and a survey on community and social infrastructure. This section provides information on the data collected in each component.

2.1. Household Questionnaire

The household questionnaire was administered to the person who best knew the business and concerns of the family, its income and expenditures, and the health of all its members. This respondent may not necessarily have been the head of the household, however the household questionnaire was not to be administered to a child. An outline of the modules contained in the household questionnaire is presented in Table 1.

Table 1: Household Questionnaire

A: Identification data
B: Household composition
C: Housing
D: Agriculture and animal husbandry
E: Expenditures
1) 7 day reference period
2) 30 day reference period
3) 3 month reference period
4) 12 month reference period
F: Income
G: Interviewer remarks

2.1.1. Identification data (module A)

This module records the *raion* and settlement (both defined in Section 3) in which the household is situated, a unique household identification number, the date of the interview and its duration, and identifies the interviewer. The unique household identification number, HID, is constructed:

$$HID = AA2 \times 1000 + AA3,$$

where AA2 is the settlement identifier and AA3 is a number ranging from 1 to the total number of households sampled from a particular settlement.

2.1.2. Household composition (module B)

This module presents a household roster which is designed to collect basic demographic information on members of the household and establishes the relationship between them. A household was defined to include people who reside in the given living quarters, share income and expenditures, and conduct housekeeping together. In determining who was in a particular household, the exact familial relationship between people was irrelevant. Children under 18, unmarried and living elsewhere as students were considered members of the

household. Children 18 years and older who were not living with the family were not considered members of the household (even if such people were materially helped by the household). In the individual questionnaires, information is only collected for household members.

2.1.3. Housing (module C)

This module collects information on the type, construction, size and ownership of the housing unit and how long the family has been residing there. It establishes the presence of amenities such as electricity, centralized heating and water supply, sanitation and telephone. Information is collected on other forms of housing owned by the family. The module also establishes the presence and approximate value of consumer durables such as refrigerators, washing machines, televisions, autos or trucks and carpets, and whether or not any of these items were sold in the last twelve months.

2.1.4. Agriculture and animal husbandry (module D)

This module establishes whether the household had the use of land for farming and animal husbandry, and if so, how much land was available and what was the ownership situation. The module considers three aspects of farming and animal husbandry:

- The home production of crops (including vegetables, fruit, grains and tobacco) over the past 12 months is recorded. There are details on the quantity produced, sold, consumed by the family and given free to relatives and others. There is no information on the value of sales of each item, but the overall value of crop production sold in the past 30 days is recorded.

- Information is recorded on the ownership of cattle, pigs, sheep, goats, horses, poultry, rabbits and bee hives. For each type of animal, the module records current numbers owned, changes in numbers over the last 12 months (and reasons for this), sales, and the respondent's estimate of the current market price of the animal.

- Details of the home production of meat, poultry, milk, eggs, honey, wool and pelts over the previous 12 months are provided. There is information on the quantity produced (and what it could currently be sold for), sold, consumed by the family and given free to relatives and others. The overall earnings from sales of animal products in the last 30 days is also recorded.

2.1.5. Expenditures (module E)

This module records household expenditures under four separate reference periods:

- *7 day reference period:* Details on the quantity purchased and amount paid for 68 foods are recorded. For important foods (bread, meat, milk, eggs, potatoes and rice) there is a record of the quantity purchased from different sources (state store, cooperative or private store). The amount spent on eating out in the past 7 days is recorded.

- *30 day reference period:* This reference period includes information on the amount spent on medicine, fuels, services (eg. public and private transportation, repair work of clothing, furniture and appliances), rental for housing and utilities (cold and hot water, heating and power). There is also data collected on miscellaneous expenses such as tuition fees, other medical treatment (excluding medicine), purchases of financial assets and other financial transactions. The value of either monetary or in-kind gifts to relatives and others is also recorded.

- *3 month reference period:* Expenditures on clothing and footwear are recorded.

- *12 month reference period:* Expenditures on household appliances, transportation, housing and furnishings are recorded.

2.1.6. Income (module F)

This module records all sources and amounts of income earned by the household over the past 30 days as well as total income earned. Where the income received was in the form of a benefit or in-kind, the household was requested to estimate the monetary value. There is information on the following income sources:

- The total amount of wage income earned by the household and also income from the sale of products from a private land plot or farm were recorded. However there is more detailed information on these income sources in the adult questionnaire and module D of the household questionnaire respectively, so this section is mainly useful for cross-checking purposes.

- Subsidies from employers and local authorities (eg. allowances for vacation, nursery school fees, food, public transport, medical treatment, housing) were recorded. The household was asked whether it received any fuel subsidies, but the value of these was not recorded.[4]

- Childcare allowances (one-time childbirth benefits, childcare benefits and single mothers' childcare benefit) were recorded and if a household was eligible for such benefits but did not receive them, then the reason for this was established.

- Gifts or charity from persons outside the household (relatives, friends, religious groups, international organizations, other organizations or private individuals) were recorded.

- Income from other sources (pensions, stipends, sickness pay, unemployment benefits, sales of the products of individual labor activity, sales of private belongings, rental property, invested capital, insurance payments, alimony payments and changes in financial assets) were recorded.

[4] Only 30 households reported receiving fuel subsidies.

2.1.7. Interviewer remarks (module G)

This section records the interviewer's opinions on the success of the interview and likely accuracy of the data collected.

2.2. Adult Questionnaire

The adult individual questionnaire was administered personally to every member of the household 14 years and older, preferably privately. Interviewers were not permitted to fill out an adult questionnaire based on answers provided by another member of the household. An outline of the modules contained in the adult questionnaire is presented inTable 2.

2.2.1. Identification data (module H)

This module records the same information provided in module A in the household questionnaire. In addition to basic demographic information, it records the household identification number, HID (=A1H3), and the position of the individual on the household roster, PID (=A1H4).

Table 2: Adult Questionnaire

H:	Identification data
I:	Migration
J:	Labor
	1) primary employment
	2) secondary employment
	3) entrepreneurial activity
	4) other work
	5) current well-being
	6) education
	7) pensioners
	8) unemployed and inactive
	9) summary variables
L:	Morbidity and use of medical facilities
M:	Self-reported health evaluation
N:	Questions for women
O:	Time use
P:	Nutrition
Q:	Anthropometric measurements
R:	Interviewer remarks

2.2.2. Migration (module I)

This module records information on:

- the birthplace of the respondent and, if applicable, the place he/she resided before moving to the current area of residence
- residential permits
- ethnicity
- language used at home and by parents
- education of parents

2.2.3. Labor (module J)

This module has information on:

- *Primary employment:* This section records information on the primary job of the respondent if it involves working in an enterprise, organization, collective or state farm, or cooperative. There are details on the respondent's occupation, primary duties, ownership of place of work, and payment in last 30 days (less deductions). There is also information on hours of work in the last 7 days and whether the respondent worked more or less than usual in the last 7 days (and reasons for which). Job satisfaction and willingness to retrain are also recorded.

- *Secondary employment:* If the respondent held an additional paid job, this section details the type of enterprise and amount paid in the last 30 days (less deductions).

- *Entrepreneurial activity:* This section records information on businesses owned (or part-owned) by the respondent. For businesses producing goods, it records what is produced and the value of finished goods and expenditures in the past 30 days. For businesses involved in trade operations, it records what was traded, whether goods were bought abroad (and from where), the value of goods sold and bought in the past 30 days and expenditures over the past 30 days. For businesses rendering services it records the type of service rendered and the value of receipts and expenditures over the past 30 days. For all types of business, there is also information on the percentage of the business owned by the respondent, who else owns the business, the value of business assets, the number of employees (both household members and not) and profit received over the past 30 days.

- *Other work:* This section records income earned from any other work other than what was mentioned in the sections above.

- *Current well-being:* The respondent is asked for his/her perception about his/her current economic situation and the prospects for the future.

- *Education:* This section measures the years of 'general secondary education' as well as completion of specialized vocational, secondary, and higher education.

- *Pensioners:* Pensioners are asked the type and amount of pension received in the last 30 days.

- *Unemployed and inactive:* This section is aimed at identifying and recording information on those who are either unemployed or who are not in the labor force. It is possible to establish the duration of unemployment as well as how long a respondent has been out of the labor force (and the reason why the respondent is not in the labor force). Discouraged jobseekers (not actively seeking work but would like to work) can be identified. For those actively seeking work, there is information on how the person has sought work, usage of the government employment service, attitude to retraining and receipt of unemployment benefits.

- *Summary questions:* Module J also provides a summary question on the total income earned in the past 30 days from all sources and the respondent's main occupation at the present.

2.2.4. Health data in the KMPS

There are five modules which collect information on health issues. The nature of the health data collected, and the way it was collected is one way in which the KMPS differs from the

usual LSMS. The following five sub-sections briefly summarize the data collected in the health related modules of the adult questionnaire.

2.2.5. Morbidity and use of medical facilities (module L)

The information in this module is provided by the respondent. The respondent was asked to describe any medical problems over the past 30 days and whether medical attention was sought. If the respondent saw a doctor in the last 30 days[5] there is information on the type of medical attention (visit to doctor or home visit), and its cost. If the respondent was hospitalized in the past 30 days, there is information on the cost of treatment, including medicine. There is also information on availability of medicine, usage and cost of preventative care, and number of days missed from work or school because of illness.

2.2.6. Self-reported health evaluation (module M)

The information in this module is provided by the respondent. The respondent was asked for his/her height, weight, and perception of health and state of mind, and ability to work and perform daily activities. The respondent was also asked whether he/she had any difficulties performing a number of activities such as walking, running, lifting, eating and dressing. For those respondents with health problems which affected their ability to perform day to day tasks, information was collected on who provided care and help. Information was collected on the existence and treatment of health problems such as diabetes, miocardial infarction and cerebral hemorrhage, and eyesight and hearing problems. Information on the respondent's usage of tea, coffee, tobacco and alcohol was also collected in this module.

2.2.7. Anthropometric measurements (module Q)

The information in this module is gathered by the interviewer. In this module data was collected on the respondents height, weight, hip and waist circumference and also whether he/she had any amputated limbs. If the interviewer did not have any medical training, then the data was collected by trained medical personnel.

2.2.8. Questions for women (module N)

The information in this module is provided by the respondent. Female respondents answered questions about their experience with pregnancy, childbirth, abortion, and birth control.

2.2.9. Nutrition (module P)

The information in this module is provided by the respondent. The respondent was asked to reconstruct from memory what food was consumed in either the preceding 24-hours or during the previous day. The interviewer asked questions to help the respondent remember what was eaten; from the answers of the respondent the interviewer assessed the type, quality and quantity of consumed food.

[5] Note that it is not explicit in the questionnaire that the visit to the doctor be related to the health problem previously described.

To help evaluate the quantity of food consumed, a 'food album' with pictures of various portions of food products and dishes (in actual size) were used[6]. In addition, quantities of foods were in units familiar to the person being questioned (for example, cups, glasses, platefuls and spoonfuls). The food albums could also be shown at the end of the questioning to help the respondent recall food which perhaps he or she had forgotten.

2.2.10. Time use (module O)

This module asked respondents to estimate time spent on different activities (and, if relevant, time spent commuting to them) over the previous seven days (not including the day of the interview). Information was collected on the following activities: working (including work at an enterprise/organization and home, entrepreneurial activity, farming, and individual labor activity); work on the garden at home, dacha or garden plot; studies; shopping for food and non-food items; obtaining household services (laundry, tailor etc.); other home duties including cooking, washing dishes and cleaning; caring for children and other relatives; sleeping and recreational activities.

2.2.11. Interviewer comments (module R)

See module G in the household questionnaire.

2.3. Child Questionnaire

The child individual questionnaire was completed for every member of the household under the age of 14 years. The questionnaire was administered to the adult member of the household who was responsible for caring for the child. An outline of the modules contained in the child questionnaire is presented in Table 3.

Modules I, O, Q and R collect information to similar their counterpart modules in the adult questionnaire. Module H collects the same information as the counterpart module in the adult questionnaire, except there is an additional variable (A1H11) which identifies the adult member of the household who answered the questions on behalf of the child.

2.3.1. Child care (module K)

[6] The technique of memory reconstruction described above and also the use of food albums are two ways in which the KMPS differs from the standard LSMS.

This module has information about the level of education of the child and if the child currently attends school, the cost of fees and textbooks are recorded. If applicable, there is information on the reason(s) why the child does not currently attend school. There is information on whether the child has missed school during the past year because of agricultural work commitments, and if applicable, how much school was missed. The module also has information on whether the child has been cared for by relatives who are not members of the household and, if so, on how many days in the last week did this occur (and the average number of hours per day). Similar questions are asked regarding those children who attended kindergarten, nursery school, or the like.

Table 3: Child Questionnaire

H: Identification data
I: Migration
K: Child care
L: Morbidity and use of medical facilities
M: Health assessment
O: Time use
P: Nutrition
Q: Anthropometric measurements
R: Interviewer comments

2.3.2. Morbidity and use of medical facilities (module L)

This module asks the same questions as its counterpart in the adult questionnaire and collects additional information vaccinations received by the child, their cost and, if applicable, reasons for not receiving them.

2.3.3. Health evaluation (module M)

This module asks the respondent the child's height and weight and for an assessment of the child's physical and mental health. Data is collected on the presence and treatment of diabetes and the presence of medical conditions such as head cold, sore throat, diarrhoea or other irregularities in defecation and leukemia. There is also information on the child's consumption of tea and coffee.

2.3.4. Nutrition (module P)

This model evaluates the food consumption of the child using the same techniques used in the nutrition module of the adult questionnaire. For those children attending school or nursery school, interviewers were instructed to additionally question the person(s) with knowledge of the child's food intake at that institution (for example, teacher, day care worker or school cafeteria worker). It should be noted that the KMPS does not contain any information on breast-feeding.

2.4. Survey of Availability and Prices of Food Products and Fuel

This survey contains three sections of information relating to retail outlets[7] selling food products in the 'local area'[8] of the households participating in the survey. The local area of the households was determined by the following method:

[7] Includes stores, kiosks, markets and street vendors. Sales of prepared food sold by stores,*kulinaria* (delicatessens), restaurants, cafes etc. were disregarded.

- housewives from the households participating in the survey were questioned, and from this a preliminary list of retail outlets was constructed.
- from this list of frequented retail outlets, a list of all streets and alleys within walking distance was constructed.
- the observer then walked down these listed streets and alleys and constructed a complete list of all retail outlets.

The survey includes preliminary identification data (from the cover page of the questionnaire) and details the *raion*, settlement identifier, census enumeration district, date of survey and also the name of the person conducting the survey. The sections contained in the survey are:[9]

· *Form A: List of all retail outlets in the neighborhood selling food, drinks and tobacco products.* This section has data on up to 20 retail outlets. There is information on location, type, hours of operation, number of employees, type of ownership, and goods sold. There are seven types of sales site: general food stores and specialized food stores selling milk and milk products; bread; meat, fish and poultry; fruits and vegetables; alcohol; tobacco products. There are three ownership classifications: state owned; non-state, cooperative, commercial etc.; and private owned.

· *Form B: List of all retail outlets in the neighborhood selling fuel:* This section has data on up to 6 retail outlets. There is information on location, type of ownership (same classifications as above) and type of fuel sold (gasoline, coal, wood, diesel fuel, kerosene).

· *Form C: Classification of retail outlets in the neighborhood selling food, drinks and tobacco products.* This section provides a table in which each trade site listed in Form A is classified by type of ownership and products sold. The purpose of this form is to help the reporter identify which trade sites are to be used in the compilation of data on product availability and prices (see form E).

---

[8] Note that this is different to the 'immediate place of residence' which is defined below in relation to the Survey of Community and Social Infrastructure. However, given that the number of observations in the Survey of Prices and Availability of Food Products and Fuel (212) is almost equal to the number of observations in the Survey of Community and Social Infrastructure (213), the two concepts must be closely related.

[9] It should be noted that the English translations of the both the survey and the instruction manual for conducting the survey are somewhat confusing. The survey is divided into 5 main sections (or 'forms') which, for simplicity, should be denoted section A through to section E. However, the English translation alternates between the original Russian letters for these sections and the English letters. In particular:
Form C in English refers to Form B in Russian (B is the third letter of the Russian alphabet). However, as the letter B is pronounced "V", "V" is used in the English translation of the instructions for the price survey.
Form D in English refers to Form G in Russian (the fourth letter of the Russian alphabet, which looks like an upper-case gamma).
Form E in English refers to Form D in Russian (the fifth letter of the Russian alphabet, which resembles a trapezoid).

- *Form D: Classification of retail outlets in the neighborhood selling fuel.* This section provides a table in which each trade site listed in Form B is classified by type of ownership and products sold. The purpose of this form is to help the reporter identify which trade sites are to be used in the compilation of data on product availability and prices.[10]

- *Form E: Availability and price of food products in different stores in the neighborhood.* This section has 14 parts each part covering a different combination of type of sales site and type of ownership.[11] Information is collected on 99 products from general grocery stores; 15 products from milk stores; 15 products from bread stores; 21 products from meat, fish and poultry stores; 26 products from vegetable and fruit stores; 8 products from alcohol stores; and 3 products from tobacco stores. The section contains information on the availability of the different products and the prices of the cheapest and most expensive types or brands of each product. The information is recorded for only *one* store in each classification.[12] This store was *not* chosen randomly; reporters were given the freedom to find the most prominent or well-stocked stores in a particular classification. For more details on the price survey sampling procedures, see Section 3.5.

2.5. Survey of Community and Social Infrastructure

Basic information on community services, infrastructure and economic structure were collected in this survey. A community, or 'immediate place of residence' is defined as the microcensus enumeration district in urban areas and the settlement (village) in rural areas. The survey includes preliminary identification data and details the*raion*, settlement identifier, microcensus enumeration district, date of survey and also the name of the person conducting the survey. In addition, there is information enabling the community data to be linked to the household data (for more details see sub-section 5.6).

The information collected on communities where sampled households live can be grouped as follows:

- Population and area. For urban communities, information was also gathered on the population and area of the entire urban area (or settlement) where the community is located.

- Rights to use of land for personal and commercial purposes.

---

[10] There was no effort to gather fuel prices in the KMPS. The reason for this was that at the time of the survey fuel prices in the Kyrgyz Republic were still centrally determined and hence there was no variation across regions. Consequently in the survey there is no form for recording product availability and prices of fuel.

[11] Privately owned stores are not covered. This form was based on that used in the Russian Longitudinal Monitoring Survey (RLMS) with adjustment to reflect the Kyrgyz diet (*buza*, *airan*, *kurt*, horse-meat and *lepeshki*, for example, were added). Note that there is a typographical error in the English translation of the form: the unit of price was soms, not roubles.

[12] In many areas some of the above classifications were not represented.

- Distance to *raion* and *oblast* centers and the nearest big city.

- Existing types of housing and types of housing available for purchase by private individuals.

- Transportation and communication infrastructure. Specifically, data were collected on: roads; telegraph, telephone, television and postal services; newspaper service; public libraries; recreational centers; and public transport.

- Presence of social service facilities such as public health facilities, schools and social welfare offices.

- Restaurants and other public eating places.

- Labor markets and Employment Service Offices. Specifically, data were collected on types of occupations available (with monthly salary), existence of Employment Service Office and whether or not any state enterprises have recently been shut down.

- Presence of services such as: banks, police, fire brigade.

- Existence of social infrastructure such as: sources of water, sanitation, electricity.

3. Sample

This section provides a brief summary of the sample and sampling frame used for the 1993 KMPS and is based on documents provided by the survey team and referenced in appendix C. The sample is designed to be fully representative of all households in the Kyrgyz Republic in the second half of 1993. Stratification was based on information on the population provided in the 1989 Census (since results from the 1994 microcensus were not available at the time of the survey).

According to the 1989 Census, there were about 856,000 families and 4,258,000 individuals living in the Kyrgyz Republic at that time (an average of about five members per family). Though the definition of 'household' used in the KMPS differs from the Census definition of 'family', this figure provided an estimate of the number of households from which the sample was to be drawn. Note that the sampling methodology assumes that any growth in the number of households since 1989 was equally distributed across regions. The target household sample size was 2,000. To allow for an estimated non-response rate of about five percent, a sample of 2,100 households was drawn. The actual number of completed household interviews was 1,938, reflecting a non response rate of 7.7 per cent. The response rate for individuals is more difficult to calculate, since some household members (eg. students under 18 studying elsewhere) could not be interviewed.

A stratified, multi-stage sampling procedure was used, with the number of stages dependent on whether households were being drawn from urban or rural areas.[13] The following is a brief description of the sampling process (summarized inTable 4).

Table 4: Stages of the sampling process

		Non self-representing strata[a]	
Stage	**Self-representing strata**	*Urban areas*	*Rural areas*
1st	microcensus enumeration districts	urban settlements (cities)	rural settlements (villages)
2nd	households	microcensus enumeration districts	household
3rd		household	

[a] - includes the two mixed urban-rural self-representing strata (see below for explanation)

3.1. The formation of strata

The Kyrgyz Republic is divided into 6 *oblasts* (see note to Table 5 for their names). These *oblasts* are further divided into 57 *raions* which fall into two broad categories: 40 county-like territories and 17 relatively large cities or sections of cities which are under the direct jurisdiction of the *oblasts* rather than the *raions* in which they are located.

A total of 21 strata were formed. These were of two types: self-representing (SR) strata (these consist of raions selected in the sample with certainty), and non-self representing (NSR) strata.

3.1.1. Self-representing strata

A total of 14 SR strata were selected. Twelve of these were cities or sections of cities which are so populous that at least some inhabitants would be expected to fall into any random sample of a given size (these are referred to as the 'urban SR strata').[14] The urban SR strata were:

- the four *raions* of the capital, Bishkek (which is also the administrative center of Chuiskaya *Oblast*);

[13] Formally, the unit of selection was the dwelling, not the household. This is because the survey team only had available a list of dwellings and, in the case of multiple households living within the same dwelling, it was generally not possible to identify the different households prior to drawing the sample. In the cases of multiple households, interviewers were given instructions on how to select one household for interviewing (these instructions are described in sub-section 3.4 below). In a few cases, interviewers had to randomly select one household to interview from the several households residing within the dwelling. However, this was so uncommon that the survey team felt justified in leaving the dwelling out of the stages outlined in Table 4. Further, when in advance of drawing the sample the survey team *was* able to identify several households living in a particular dwelling, the households were listed separately before using systematic sampling. Thus, the survey is not unambiguously a sample of dwellings either.

[14] Note that a city which is the administrative center (or capital) of an oblast was classified as self-representing even if it did not have a large population.

- the five other *oblast* administrative centers (each consisting of one *raion*): Dzhelal-Abad; Naryn; Talass; Osh; Balykchi;
- three other major cities (each consisting of one *raion*): Karakol (formerly Przheval'sk); Tokmak, and Kara-Balta.

The other two SR strata were the two *raions* Suzakskii and Kara-Suiskii which were selected with certainty for reasons outlined below (these are referred to as the two 'mixed urban-rural SR strata').

3.1.2. Non self-representing strata

Forty-five *raions* remained on the list after the selection of the 12 urban SR strata. Forty of these were territories *raions* and five were cities under the direct jurisdiction of the *oblast* in which they were located. The five cities (Uzgen, Tash-Kumyr, Kyzyl-Kiya, Kara-Kul', and Mali-Sai) were combined with the territories in which they are geographically situated, thus increasing the heterogeneity of those *raions*. The second group of NSR strata was therefore selected from forty *raions* (some of which were combined with the five cities mentioned above).

Table 5: Distribution of households in self-representing strata

Dzhelal-Abad (DA)	39
Karakol (IK)	38
Balykchi (IK)	26
Osh (OSH)	112
Naryn (N)	20
Talass (T)	19
Tokmak (CHU)	55
Kara-Balta (CHU)	38
Bishkek (CHU)	
Leninskii *raion*	79
Oktibrskii *raion*	100
Pervomaiskii *raion*	84
Sverlovskii *raion*	108
Total households selected from SR strata	718

Note: Abbreviations for *oblasts* are: Dzalal-Abad (DA), Issyk-Kul'skaya (IK), Osh (OSH), Naryinskaya (N), Talasskaya (T), Chuiskaya (CHU). The table does not include the two mixed urban-rural SR stratawhich are included in Table 6.

The NSR strata were identified on the basis of three characteristics: geographical conditions (mountains, valleys or a mix of the two); type of production (agriculture, industry or a mix of the two); and ethnic composition (Kyrgyz, mostly Kyrgyz and Uzbek; or mostly Kyrgyz and Russian-speaking). Of the 27 possible strata, six were formed:

I. mountains; agriculture and animal husbandry; predominantly Kyrgyz population.

II. mountains; agriculture, animal husbandry and nurseries; predominantly Kyrgyz population.

III. mountains; agriculture-industry, predominantly Kyrgyz and Uzbek population.

IV. valleys; agriculture; predominantly Kyrgyz and Russian-speaking population.

V. valleys and mountains; agriculture, predominantly Kyrgyz and with Uzbek population.

VI. valleys, agriculture-industry, predominantly Kyrgyz and Russian-speaking population.

Table 6: Distribution of households in non self-representing strata

	Population	Sample	U	R
Stratum 1 (189)				
Toguz-Torouzskii (DA)	3,598			
Chatkal'skii (DA)	3,705			
Ak-Talinskii (N)	5,063			
Tian'-Shan'skii (N)	6,481			
At-Bashinskii (N)	8,159	94	16*	78
Tonskii (IK)	10,233			
Alaiskii (OSH)	11,357			
Dzheti-Oguzskii (IK)	14,083	95	16*	79
Toktogul'skii (DA) + Kara-Kul'	17,901			
Stratum 2 (166)				
Manasskii (T)	5,054			
Kara-Burinskii (Kirovskii) (T)	7,878			
Bakai-Atinskii Leninpol'skii (T)	8,584	83	2*	81
Talasskii (T)	9,104			
Kochkorskii (N)	9,295			
Dzhumgal'skii (N)	9,360			
Batkenskii (OSH)	9,457	83	2*	81
Kara-Kul'dzhinskii Sovietskii (OSH)	10,496			
Stratum 3 (153)				
Tiupskii (IK)	11,066			
Ak-Suiskii (IK)	11,487	76	7*	69
Aksyiskii (DA) + Tash-Kumyr	18,331			
Naukatskii (OSH)	24,440	77	8	69
Stratum 4 (213)				
Chuiskii (CHU)	10,459			
Panfilovskii (CHU)	12,734	106	13	93
Issyk-Kul'skii (IK) + Cholpon Ata	14,702			
Moskovskii (CHU)	20,739			
Sokulukskii (CHU)	32,504	107	13	94
Stratum 5 (208)				
Kalininskii (CHU)	11,885			
Issyk-Atinskii (CHU)	12,356	104	22	82
Keminskii (CHU	16,085			
Kantskii (CHU)	20,043			
Adamedivskii (CHU)	28,371	104	22*	82
Stratum 6 (140)				
Missing Name				
Aravanskii (OSH)	12,070	70	13	57
Bazar-Korgonskii (DA)	14,926			
Nookentskii (Leninskii) (DA) + Mali Sai	23,438	70	13*	57
Strata 7a and 7b (148)				
Suzakskii (DA) + Kok-Yangak	25,501	74	9	65
Kara-Suiskii (OSH) + Kara-Suu	37,782	74	8	66
Stratum 8 (168)				
Liailiakskii (OSH) + Suliukta	15,980	84	27	57
Uzgenskii (OSH) + Uzgen	27,365	84	27	57
Kadamzhaiskii (Frunzenskii) (OSH) + Kyzyl-Kia	28,100			
Total households selected from NSR strata		1,385	218	1,167

Note: Numbers in parentheses following stratum titles are the numbers of households sampled from that stratum. The population figures are number of families and come from the 1989 Census. The column titles 'U' and 'R' denote 'urban' and 'rural' respectively. The table includes strata 7a and 7b which technically are SR strata. See text for explanation of asterisks.

Based on the 1989 Census, the household populations of strata II and V were about twice large as the household populations of the other strata. To ensure that all strata were proportionally represented, strata II and V were therefore both split into two, resulting in a total of eight strata (henceforth named using arabic numerals so as to distinguish them from the above). The survey team envisioned that stratum 7 would be a NSR stratum. However, as there were only two *raions* (Suzakskii and Kara-Suiskii) in this stratum, both of which were therefore chosen with certainty. Therefore stratum 7 technically became two separate SR strata (7a and 7b), with each strata containing a single *raion* (these are referred to as the mixed urban-rural SR strata). Although these *raions* technically were SR strata, they were treated in the sampling process as if they were NSR strata (for example, in the method that households were selected from them). More details on this are presented below. The NSR strata (and also the mixed urban-rural SR strata) are listed in Table 6.

3.2. The selection of primary sampling units

The nature of the primary sampling units (PSU) differed according to whether they came from SR or NSR strata. In the urban SR strata the PSU were microcensus 'enumeration districts' (ED).[15] Based on the 1989 Census, each microcensus ED was expected to contain about 414 individuals (less than 100 households). It was considered appropriate to choose eight to ten households from a given microcensus ED and therefore enough were selected to yield the desired number of urban households from the particular stratum. The districts were chosen with equal probability and no substitution was permitted.[16]

In the seven NSR strata, the PSU were *raions*. Two were selected from each stratum with probability proportional to size (PPS), as measured by reported households in the 1989 Census. As mentioned above, strata 7a and 7b were treated in the sampling process as NSR strata. In this sense their PSU were the *raions* themselves.

3.3. The selection of secondary sampling units

The selection of secondary sampling units (SSU) differed depending on whether the PSU was drawn from a SR or NSR stratum.

3.3.1. SSU within selected PSU from urban SR strata

The SSU selected from microcensus ED in the 12 urban SR strata were the households (these were also the last stage sampling units for these strata).

3.3.2. SSU within selected PSU from NSR strata

---

[15] These districts were made available through the cooperation of GOSKOMSTAT of the Kyrgyz Republic in connection with its preparation for a microcensus scheduled for early 1994. GOSKOMSTAT no longer had complete records of the boundaries of the ED used in the 1989 Census.

[16] There was not enough information to use probability proportional to size in the selection of microcensus ED, however the size in terms of population tends to be roughly equivalent across districts.

Within the *raions* selected as PSU from NSR strata (and also the mixed urban-rural SR strata), 'settlements' (or areas where people are living) were classified as *gorodskoi* (urban) or rural.[17] The number of urban settlements within a *raion* generally did not exceed two or three.

SSU selected from urban settlements

It should be emphasized that urban settlements were *not* the SSU; urban settlements were selected, and then the SSU were selected from these settlements. If there was only one urban settlement in a *raion*, then it was selected. If there was more than one urban settlement, then one was selected using PPS for each 15 urban households required from the *raion*. There were seven *raions* selected as PSU from NSR strata that did not contain urban settlements, even though they represented strata in which there *were* urban settlements (these *raions* are indicated in Table 6 by the asterisks against the target number of urban households to be sampled). This problem (involving 78 of the 2,100 target households) arose because the number of urban households to be sampled from a particular stratum was calculated taking into account the Census information on household populations in all of the *raions* within the stratum, not just those that were selected as PSU.[18]

The problem was rectified using substitution. In strata 3, 5, and 6, only one of the two selected *raions* had urban settlements. Consequently all of the required urban settlements were drawn from that particular *raion*. For example, in stratum 3 the 7 households which were to be drawn from Ak-Suiskii *raion* were instead drawn from Naukatskii *raion*. In stratum 1, *both* selected *raions* did not contain urban settlements, the 32 target urban households were drawn from urban settlements in Toktogul'skii *raion*. In stratum 2, where both selected *raions* did not contain urban settlements, the 4 target households were drawn from the city Talass, which was already in the sample as a SR stratum.

The SSU selected from urban settlements were microcensus ED. The microcensus ED were selected in the same manner as described in sub-section 3.2 above.

SSU selected from rural settlements

In rural areas, villages were the SSUs. Effort was made to ensure that the ethnic composition of villages was properly represented in the sample. Within each selected PSU, data from the 1989 Census was used to group villages by ethnicity. For example, in stratum 1 (in 1989), 94.5 per cent of villages were Kyrgyz; 3 per cent were Russian; 0.8 per cent

[17] In rural areas, settlements are villages, while in urban areas settlements are towns, cities or 'villages of the city type' ('PGTs' in Russian). The documentation provided by the survey team uses the term 'population point' in place of settlement. It should be noted that the term 'urban' is a somewhat misleading translation of *gorodskoi*, since many settlements counted in official statistics as cities or as PGTs have populations of only 5,000 to 15,000 individuals. Unincorporated towns which are larger than some PGTs are, however, counted as rural.

[18] This problem arose because the sample was conducted using relatively large units (raions) as PSU and the survey team did not have sufficient time or data to construct PSU in which both urban and rural settlements could always be found.

were Uzbek; 1.7 per cent were other. The number of households chosen from each group of villages was made proportional to the number of villages of each type in the stratum. The selection of individual villages from the ethnic groupings was random and no more than 18 to 20 households were selected from a given village. Note that quotas were not used, so the exact distribution of households by ethnicity was not guaranteed at this level.

3.4. The selection of last stage sampling units (households)

For all strata, the last stage sampling unit was the household[19], with the households being drawn randomly from the selected SSUs. Interviewers were given a list of addresses and the names of the person responsible for each dwelling (akin to a lessee). The interviewer was obliged to interview the household of that person at that address. If the person responsible for the apartment did not reside there, but relatives did live there, the interviewer was obliged to interview the related household at the address. If no related household lived there, the interviewer was obliged to interview whatever household did live there. If more than one household occupied the dwelling, and that fact was not registered before selection, the interviewer was obliged to randomly select one household.

According to the 1989 Census, the 12 SR urban strata contained 34.2 per cent of households, and therefore 718 of the 2,100 households were drawn from them. The number of households drawn from each SR urban strata was proportional to its total population of households (see Table 5).

The remaining 1,385 of the target 2,100 households were selected from the NSR strata and also the two mixed urban-rural SR strata. The number of households selected from each of these strata was proportional to the total population of households within the stratum, and is indicated in parentheses beside the stratum title in Table 6. However, the number of households to be surveyed within each stratum was divided equally between the two *raions* selected as PSUs, even though the population of the *raions* differed.[20]

3.5. Price survey sample procedures

This sub-section provides an outline of the sampling procedures used in price survey (for details on the data collected in the price survey, see sub-section 2.4). For more details on the sampling procedure, refer to the document titled 'Instructions for the Price Survey' referenced in appendix C.

---

[19] For the definition of 'household' see sub-section 2.1.

[20] When stratum 7 became the two mixed urban-rural SR strata (7a and 7b), the target sample size should have been distributed between the two raions using PPS rather than equally (this would have been consistent with the distribution of the target households across the 12 urban SR strata). In other words, 14 fewer households should have been drawn from the Suzakskii raion and 14 more from the Kara-Suiskii raion. Weights could be applied to correct for this mistake, however the survey team felt that this would make no appreciable difference to analysis.

The price survey was designed to provide a relatively inexpensive method of measuring the prices to which the households *in the sample* were exposed. It was designed to provide covariates (contextual variables) for the household survey. The sample was not weighted to represent the price of a foodbasket in the Kyrgyz Republic.

If there existed more than store within each classification of 'potential' stores (see Form C in the Survey), the reporter was required to select the store at which people would be most likely to shop. This was not determined rigorously; however since the sampling points were normally quite small in area, it was not difficult for the reporter to build an impression. As the sample points in many cases were either quite small or were rural, often there did not exist stores of certain classifications. If in a particular sample point there did not exist a store in a key classification (marked by an asterisk in Form C), the reporter was obliged to search for such a store within an area one kilometer of the borders of the sample point.

3.6. Survey as implemented

As mentioned above, the sample was designed to be fully representative and therefore self-weighting. The document on the sampling procedure referenced in appendix C contains information on post-sampling representativeness and also design effects and the efficiency of the sample.

4. Field Work

This section provides a brief description of how the field work was conducted, based on information contained in the documents prepared by the survey team, referenced in appendix C.

4.1. Design, translation, piloting and printing of the questionnaires

Given that there was a need to provide survey data in a relatively short period of time, it was decided to base the KMPS questionnaires on the questionnaires designed for the Russian Longitudinal Monitoring Survey. However, fairly substantial changes were made to the household and individual questionnaires to ensure accurate measurement of living standards in the Kyrgyz Republic.

The individual and household questionnaires were translated from Russian into Kyrgyz in August 1993. The child and adult questionnaires, though quite similar, were translated independently, thus allowing a comparison of translations to check for mistakes. The questionnaires were also reviewed in Osh by a sociologist. As the price and community questionnaires were to be filled out by professional staff members, there was no need to translate them from Russian. The individual and household questionnaires were also translated into Uzbek, for use in the territory around Osh (which borders Uzbekistan). All

the translations were completed by late September 1993. Note that the English version is therefore a translation of the original.

In August the individual and household questionnaires were administered to approximately 100 households in Bishkek and its rural suburbs. This process revealed several branching errors, but resulted in no major changes to the questionnaires. The branching errors and other, small, errors were fixed and the final versions of the individual and household questionnaires were printed in Bishkek. The price and community questionnaires, as well as the instruction booklets and the food albums were printed in Moscow.

4.2. Organization of field work

Sixteen people were sent from Moscow to the Kyrgyz Republic in early October 1993 to conduct interviewer training and to organize the survey. Eight of these people were sociologists from the Institute of Sociology of the Russian Academy of Sciences and the other eight were dietitians from the Institute of Nutrition of the Russian Academy of Sciences.

Teams were dispatched to five major cities: Dzhalal-Abad, Talass, Naryn, Karakul, and Osh. Each team consisted of at least three people: a supervisor (sociologist) from Moscow, a dietitian from Moscow, and a sociologist representative from the Republican Center for the Study of Public Opinion which is headquartered in Bishkek. Double-sized contingents were sent to Dzhalal-Abad and Talass because more households were sampled near those centers. The survey team invited GOSKOMSTAT (Bishkek office) to send representatives to observe the sampling process; two were sent to Dzhalal-Abad and Karakul. In each location, the supervisor (sociologist) from Moscow was responsible for finalizing the sample of households. Each team was also responsible for training interviewers and for organizing the local staff.

To summarize, the field work was organized around the following positions:

1. supervisor (sociologist) from either Moscow or Bishkek;

2. dietitian and interviewer trainer, both of whom trained interviewers and checked their initial work;

3. local supervisor employed by the Republican Center for the Study of Public Opinion, who worked with the sociologist from Bishkek in organizing local workers;

4. local brigade leaders, who conducted the price surveys and checked the work of interviewers;

5. doctors and nurses hired to perform the measurements as part of the health evaluation sections of the individual questionnaires;

6. interviewers, responsible for locating respondents, scheduling household and individual interviews, and returning questionnaires to administrators;

7. inspectors sent from Bishkek to confirm that the proper families were interviewed, that the questionnaire had all been administered, and that families had been paid.[21]

4.3. Interviewer training

A total of 178 interviewers were recruited, of which 65 percent had a higher education and another 22 percent had specialized secondary education (akin to junior college). The majority of interviewers were professionals[22]; only nine graduate students and six other students participated as interviewers. All interviewers were thoroughly trained with the training also serving as a screening device. Several trainees were dismissed before field work began because they were perceived as unsuitable for the task.

The following is a summary of the training provided:

1. Interviewers were lectured on the general principles of face-to-face interviewing. A 70-minute video tape entitled "Introduction to Interviewing" was shown.

2. Interviewers were required to read through the entire questionnaire in advance, then to fill out the questionnaire themselves.

3. Interviewers were shown an example of a good interview with commentary, again using a video tape. The tape included a section on the diet.

4. Interviewers were introduced to the interviewer instruction manuals.

5. Interviewers practice interviewing in groups of three. One assumed the role of interviewer; another, the role of respondent; the third observed to see that the interview was conducted properly. The trainer and perhaps some other experienced interviewers circulated among the groups to observe proceedings.

6. Interviewers were given written exercises which tested their ability to react properly to certain difficult situations in administering the questionnaire.

7. Interviewers reviewed the administrative procedures pertaining to the survey (outlined in the interviewer instruction manuals).

8. Interviewers practiced persuading reluctant respondents to participate in the survey.

9. Interviewers were required to complete at least one practice interview with a household that was not in the sample (and preferably not containing relatives).

---

[21] Note that it is not standard LSMS practice to pay participating families.
[22] For example, doctors, nurses, teachers, engineers, and bookkeepers.

10. Interviewers were observed during their first three interviews (or until they had demonstrated competency).

4.4. Field work

In most cases, the chartered bus which took the teams from Bishkek to the five major cities was used to take interviewers to surrounding rural areas. In other cases, cars were hired. In urban areas, public transportation was used. As all information had to be direct (i.e. one person could not answer the individual questionnaires for all household members), the interviewer was usually obliged to return to the household more than once.[23] Furthermore, unless the interviewer had medical credentials, a nurse or doctor had to visit to get the necessary measurements for the health evaluation section of the individual questionnaires. In all locations, field work began by October 18. In some areas with small samples, field work ended by the first week of November. Interviewing was finished in all areas by the first week of December.

4.5. Editing, coding, data entry and cleaning

The local supervisors were required to examine the questionnaires to locate problems which could be remedied in the field. Such problems included missing key demographic information and problem with household and individual identification numbers. All questionnaires were then sent to Bishkek, where they were again checked for identification number problems and then to Moscow, where yet another ID check was performed.

Open-ended questions (eg. occupation and nationality questions) were not immediately coded. Instead, the responses were entered into the data set in text, to be coded at a later date. Codes for all open-ended questions except occupation were made available in mid-February. Occupation codes were made available in June 1994.

Data entry and verification of the household questionnaires was completed by a private data entry firm by January 25. All other data entry was handled in-house using the SPSS data program. The first entry of the 10,000 child and adult questionnaires began on December 20, 1993; the verification pass began on January 20 and was completed by February 2. Entry of the community and price surveys began in late January and was completed in two weeks.

5. Using the Data

This section details some aspects of the data that users should keep in mind. For several reasons, it is strongly recommended that researchers refer to the questionnaire while using the data. First, most of the codes are contained in the questionnaire itself (for details on

[23] Variables measuring interview duration are included in module A of the household questionnaire and module H of the adult and child questionnaires.

codes not contained in the questionnaire, see appendix F). Second, and more importantly, extensive use is made of *skip patterns*. These were used so as to maximize the ease with which the interview was conducted by ensuring that questions which were not relevant to a particular household or respondent were not asked. The researcher must be aware of these skip patterns so that the data are properly interpreted. In most cases, the skip instructions are clearly written next to a particular response code. If there is no instruction, then the next question should have been asked regardless of the response.

5.1. Data set and variable names

Appendix B lists the raw data sets which are available. This sub-section briefly describes the structure of the variables names within each data set.

All variables within the household data sets, KHHLD and KINDIVH, are prefixed with the letter 'A'. The second letter in the variable name denotes the relevant module. Thus, for example, AC1 is the first question of the third module and AF1 is the first question of the sixth module. In the adult data set (KADULT) and the child data set (KCHILD), all variables are prefixed with 'A1'. As with variables in the household data set, the next letter in the variable name is the relevant module (for example, A1I1 is the first question of the second module in the adult questionnaire). The nutrition data for both adults and children (module P) are found in separate data sets, named KADIET and KCHDIET respectively.

Data from the Survey of Availability and Prices of Food Products and Fuel are found in the data sets KPRICE1 to KPRICE3. All variable names in the survey are prefixed by the letters 'AY'. The variables containing the preliminary identification data also contain the letters 'COV' (for coverpage) in the variable name. For example, AYCOV2 is the variable defining the settlement in which the survey was conducted. For the remainder of the survey, the variable names are prefixed by 'AY' plus a letter signifying the particular section or 'form' from which the variable comes. For example, the variable names for the retail outlet which is listed first in form A are prefixed by 'AYA1' and the variable names for the retail outlet which is listed first in form B are prefixed by 'AYB1'.

Data from the Survey of Community and Social Infrastructure are found in the data set KCOMM. All variables in the survey are prefixed by 'A1X'. The variables containing the preliminary survey details also contain the letter 'C' in the variable name. For example, A1XC2 is the variable defining the settlement in which the survey was conducted.

5.2. Using the nutrition data

The adult nutrition data (module P) are in the data set KADIET and the child nutrition data are in the data set KCHDIET. There is one observation for every meal consumed by the individual.[24] Each observation has information on the quantity of each item of food

[24] On average, there are 3 observations (meals) per adult and 4 observations per child.

consumed during the meal (A1P1_X_6, where 'X' is a number with range 1-10) and a numeric code identifying the food item (A1P1_X_7).[25]

5.3. Data quality

There was a fairly short time period for survey design, field work and provision the data. This sub-section provides some details on the general quality of the data.

5.3.1. Data cleaning

After double-entry verification, outlier values were flagged by the cleaning program, and the original questionnaires were examined to determine whether there was a mistake. If there was some basis for changing the values (for example, it was obvious than an answer was recorded in grams rather than kilograms), the value was corrected. However, at the request of the World Bank, outliers were left in the data set (unless there was specific evidence of a mistake). Thus, the researcher is responsible for defining outliers and deciding how to treat them (see sub-section 6.4 below for discussion of outliers in the expenditure data).

5.3.2. Missing value codes

There are three missing value codes provided for each variable. For each variable, the number of digits in the missing value code will be equal to the maximum number of digits of a legitimate response for that question. This ensures that for continuous variables, the missing value code is not confused for a legitimate response. The missing value codes employed for each variable are:

- 'don't know' - the highest possible code ending in '7';
- 'refused' - the highest possible code ending in '8';
- 'missing' - the highest possible code ending in '9'.

For example, for all variables occupying fields of two columns, the three missing variable codes are 97, 98 and 99 respectively. For all variables occupying fields of five columns, the three missing value codes are 99997, 999998 and 99999 respectively.[26]

---

[25] The food codes are used to provide nutrition information for each food item (for example, energy and protein content).

[26] Note that there was some difficulty in the data entry of long missing value codes. For example, in an eight-digit variable, a data-entry operator may have omitted one nine in the missing value code and entered the incorrect seven digit "don't know" code 9999997 instead of the proper eight-digit code, 99999997. Such a mistake could seriously affect data analysis since a seven-digit code would not be treated as a missing value if it was assumed that the code had eight digits, not seven. Another potential problem with the coding of missing values is the situation where missing value codes of differing lengths were entered for the*same variable*. An example of this is the variable measuring interview duration (hours) for the adult questionnaire, A1H8_1. There are 78 cases of a '9' being entered for this variable, and yet there are also two cases where a '99' has been entered. Normally, such mistakes were caught in the process of double-entry verification, however the researcher should be aware of these potential problems.

It should be noted that the missing value code which ends in '9' was not printed in the questionnaires since it was used only in those situations when a codeable response was absent because of interviewer error. That is, a missing value code which ends in '9' is to be distinguished from a legitimate skip which occurred when, based on a previous response of the respondent, a particular question was not asked of the respondent. For example, when a household states it did not purchase a particular food, the expenditure and quantity questions are skipped. A legitimate skip was coded as a '.'.

5.4. Specific data issues

This sub-section provides specific details relevant to using the data in the KMPS.[27] To this date, the majority of the research using the KMPS has focused on the individual and household questionnaires and consequently, there is no specific information on the adequacy of the data in the survey of availability and prices of food products and fuel and the survey of community and social infrastructure

5.4.1. Children's nutrition and health data

Users should be aware of two aspects of the data relevant to assessing the nutrition and health status of children in the Kyrgyz Republic. First, the age of young children is reported only to the nearest year, rather than the nearest month. Hence, it is not possible to use the anthropometric data to compute height or weight for age according to international norms. Second, as mentioned in sub-section 2.3, there is no information on breast-feeding.

5.4.2. Seasonality of the data

As the interviews were conducted just after the major harvest time, the production figures for agricultural and animal husbandry will be higher than at other times during the year. Also, the estimates of expenditure on heating will similarly be lower because of seasonal factors.

5.4.3. Non-response of individuals

An individual questionnaire was not completed for every household member, with the overall non-response rate for individuals being 5 percent. The non-response rate varied substantially between regions, with Narunskaya *oblast* having the highest non-response rate of 14.6 percent and Talasskaya *oblast* having a non-response rate of only 0.4 percent.

5.4.4. Labor data

A problem with the adult questionnaire is that there is no question asking for hours worked at additional job (although question A1J33 records earnings at additional job). This may not be an important omission as, according to question A1J29, only 2.1 per cent of those

[27] The contributions of Jane Falkingham and Raylynn Oliver were very useful in the writing of this sub-section.

working in an enterprise, collective farm or state farm (A1J1=1) are working such additional paid jobs. Another aspect of the labor data that should be taken into account by researchers is the fact that question A1J100 asks unemployed or inactive individuals whether they have tried to find work in the last 30 days. Since the standard definition of unemployment (used by, for example, the ILO) includes people looking for work in the last seven days, the above use of a 30 day reference period may lead to overestimation of the unemployed and underestimation of the inactive compared with this standard method.

5.4.5. Typographical errors in the English version of the survey

The following typographical errors have been identified in the English version of the household questionnaire:

Page	*Listed in the questionnaire*	*Variable name in data set*
19	AB111	AD34
27	AD105	AD106
27	AD106	AD105
33	AE2.1A to AE9.1C	AE1.2A to AE1.9C

5.4.6. Warning on the use of missing value codes

The following problems with missing value codes in the household questionnaire have been identified:

- for variable AC20, the missing value codes are 999997, 999998, and 999999, but there are valid answers higher than 1,000,000.
- for variable AC41, 95 and 96 are also missing value codes.
- for variable AD141_2E the missing value codes are 99997, 99998, and 99999 even though for all other variables in that section the missing value codes are 999997, 999998, and 999999.
- for variable A1J41, 999996 is also a missing value code.
- for the household with HID=51014 the missing value code for AF14_2B is 999997.02.

5.5. Weighting factors

As discussed in Section 3, the sample was designed to be self-weighting and therefore there are no weighting variables.

5.6. Linking components of the data

This sub-section describes the linkages between various components of the KMPS. Some of the variables referred to are described fully in Section 2 above.

5.6.1. Individuals to households

Adults and children can be linked to the households to which they belong via the household identification number (HID).

5.6.2. Children to adults

There does not appear to be an easy way of linking children to their parents. However, it can be done using information from the household roster. The variable A1H11 in the child questionnaire identifies the adult member of the household (who may or may not be the parent) who answered the questions on behalf of the child.

5.6.3. Household to community level data

The variable COMMID in both the household data set and the survey of community and social infrastructure data set can be used to link households to their relevant community level data.

5.6.4. Household to relevant data on availability and prices of food products and fuel

The variable COMMID in both the household data set and the survey on the availability and prices of food products and fuel data set can be used to link households to prices in their relevant local area.[28]

5.6.5. Nutrition data to other individual data and to household data

The nutrition data in KADIET (KCHDIET) can be linked to the other individual level data in KADULT (KCHILD) via the variables HID and PID. The nutrition data can be linked to household data via the variable HID on the nutrition data sets and HID on KHHLD.

6. Constructed Data Sets

Analysis of the KMPS data has made use of constructed variables pertaining to both the household and the individual. These variables are available to the public on the condition that the World Bank and its staff are under no obligation to provide further services to users. The construction of the variables reflects methodological choices of the individual researchers and some users may prefer to calculate their own versions of these variables, based on the original data. Appendix B lists the constructed data sets which are available. This section describes the construction of the variables.[29]

[28] As noted in sub-section 2.4 above, 'local area' is closely related to 'immediate place of residence' or community as defined in sub-section 2.5. It should be noted that there are some problems with merging records from the community, price and household datasets. In particular, there are 26 cases where a particular value of COMMID appears in one of the three datasets, but does not appear in either one or both of the other two datasets.

[29] All variables are continuous, unless the categories are explicitly mentioned. The names of the variables included on the constructed files are in bold text. Some of the variables in the adult and household

6.1. Household demographic and relationship variables

The file KINDIV contains basic demographic information on the household members and also variables describing the relationships between them. The variables were constructed from the data in the household roster (module B of the Household Questionnaire), and there is one record per household member. The variables are listed inTable 7 and a brief description is presented below.

The variable **HID** is the household identification number and **PID** is the position of the individual on the household roster.

Table 7: Constructed variables: household demographic and relationship variables

HID	household identification number
PID	position on household roster
REL_X	relationship to household member in Xth position on household roster (where X ranges 01-15)
MTHHOME	number of months in home (of last 12)
NGHTHOME	number of nights in home (of last 7)
SEX	gender
AGE	age in years
MARITAL	marital status
INTVIEW	individual questionnaire administered?
NAME	name of individual

The variables **REL_X** (where X ranges from 01-15) describe the relationship of the individual to the other members of the household. The categories of REL_X are:

0) self
1) husband/wife
2) father/mother
3) stepfather/stepmother
4) son/daughter
5) stepson/stepdaughter
6) brother/sister
7) stepbrother/stepsister
8) grandfather/grandmother
9) grandson/grand-daughter
10) nephew/niece
11) father-in-law/mother-in-law
12) brother-in-law/sister-in-law/son-in-law/daughter-in-law
13) other relative
14) other relationship (not familial)

MTHHOME is the number of months (out of the last year) the individual spent living in the home. **NGHTHOME** is the number of nights (in the last week) the individual spent sleeping in the home. **SEX** is equal to one if the individual is male and two if the individual is female. **AGE** is the age of the individual in years. **MARITAL** describes the marital status of the individual and has the categories:

---

characteristics constructed files are derived from variables created by the survey team. For details on the structure of the variable names used in the KMPS, see sub-section 5.1.

1) never married
2) married
3) divorced and not married
4) widow (widower)
7) don't know
8) refused
9) missing

INTVIEW is a variable equal to one if an individual questionnaire was administered to the person, and two otherwise.

6.2. Adult characteristics

The file CONADULT contains constructed variables describing the characteristics of adults in the sample. The variables were constructed from data in the adult survey. The variables are listed in Table 8 and a brief description is presented below.

6.2.1. Identification variables

PID is the position of the individual on the household roster and it is the variable of the same name in the adult data set. **HID** is the identification number of the individual's household.

6.2.2. Demographic characteristics

AGE is age in years and is equal to 1993 minus year of birth (A1H6). **GENDER** is the variable A1H5 and is equal to one if the individual is male and two if the individual is female. Ethnicity is summarized by the variable**ETHNC** which is derived from A1I15 and has the categories:[30]

1) Kyrgyz (A1I15=1)
2) Russian (A1I15=2)
3) Other Slavic(A1I15=4,7,25,39,17,18,32, or 37)
4) Uzbek (A1I15=3)
5) Other

HHEAD is equal to one if the individual is the head of his or her household and zero otherwise.[31]

6.2.3. Education data

[30] See appendix F for the codes of A1I15.

[31] The household head was determined as follows: the oldest male aged 18-59; if there were no males aged 18-54 years, then the oldest female aged 18-54 years; if there were no females aged 18-54 years, then the youngest male aged 60 years and over; if there were no males aged 18 years and over, then the youngest female aged 55 years and over; if there were no females aged 18 years and over, then the oldest person.

The adult survey contains information on years of 'general secondary school' (A1J84) but not total years of education.[32] The education level of the individual is constructed as a categorical variable, based on responses to several of the questions in the survey. Questions A1J85 and A1J86 ask whether the respondent had completed studies other than primary and secondary education and if so, what studies were completed. These extra studies are classified as[33]:

1. Vocational courses;
2. Vocational-technical school (not granting secondary diploma);
3. Vocational-technical school (with secondary education);
4. Tekhnikum, medical, music school, school of education;
5. Institute, university, academy;
6. Graduate school, residency.

Table 8: Constructed variables: adult characteristics

PID	position on household roster
HID	household identification number
AGE	age in years
GENDER	gender
ETHNC	ethnic group
HHEAD	household head status
EDLEVEL	education level
ECSTAT	economic status
LFSTAT	labor force status
OCCUP	occupation
OCCUPC	occupation group

The constructed variable **EDLEVEL** has the following eight categories:[34]

1) *primary education only; no formal education:* less than 8 years primary and secondary schooling;
2) *did not complete high school:* 8-9 years of primary and secondary schooling;
3) *high school only:* 10 or more years of primary and secondary schooling and did not study elsewhere;
4) *high school and 'other':* 10 or more years of primary and secondary schooling and completed one (or more) of the first three fields of extra study listed above;
5) *higher education:* 10 or more years of primary and secondary schooling and completed one (or more) of the second three fields of extra study listed above.

6.2.4. Economic variables

[32] An understanding of the education system in the Kyrgyz Republic is complicated by the coexistence of three distinct schooling systems. Individuals can earn a diploma in 'general secondary school' by completing grades 1-10, or more recently, 1-11. This diploma itself does not provide large economic returns, however it can lead to higher education. It is also possible to leave general secondary school after grade 8 and spend two years in a vocational school, a path which apparently provides good economic returns (relative to a diploma without further education), but is considered to have low in prestige. A third education path is to leave general secondary school before obtaining a diploma and then obtain training in specialized schools. Because of the different possible combinations of general secondary schooling and other studies, the years of general secondary schooling variable (A1J84) alone is not a particularly good indicator of human capital and expected economic returns.

[33] The classifications appear to be distinct and ordered increasing in human capital 'content'.

[34] Note that there were some problems with the reliability of the education data. Of the working age population, 24 people could not be assigned an education level because of missing data (these individuals have a '.' recorded for EDLEVEL). Also there were 4 cases of individuals reporting that they had less than 8 years primary and secondary education, yet also having completed university or graduate studies. It was decided to not classify these individuals as higher educated.

The variable **LFSTAT** describes the labor force status of individuals of working age (16-60 years for men, and 16-55 years for women).[35] LFSTAT takes the following values: 1 if the person was employed, 2 if the person was unemployed, and 3 if the person was not in the labor force.

Labor force status was ascertained using information from the following questions in the labor module of the adult questionnaire: A1J1, A1J91, A1J99, A1J100, A1J104 and A1J112. While A1J112 was designed to be a summary variable describing the respondent's current economic situation, there were some inconsistencies between answers to this question and the others.[36] It was decided that more accurate description of labor force status would be obtained by using information from a range of questions, rather than just relying on A1J112 alone. An attempt was made to make LFSTAT conform with Western notions of labor force status, however in some cases the information contained in the survey did not facilitate a completely comparable measure of a person's labor force status.

The *employed* refers to those employed as hired labor in an enterprise, organization, collective or state farm, or cooperative. Those on temporary official leave for health or maternity reasons are included. It also includes those who are self-employed (engaged in individual labor activity or who are farmers or entrepreneurs). It does not include housewives or full-time students, unless they are engaged in part-time work.

The *unemployed* refers to those who are not employed by an enterprise etc., who are not self-employed, who are not full-time students; and who have attempted to find work during the last 30 days. Persons looking for work but, for health reasons physically unable to work at present, are not classified as unemployed. The unemployed also includes those who are not unemployed using the above definition, but who are registered with the state employment service and have unemployed status. Note that since the standard definition of unemployment includes people looking for work in the last seven days, the above use of a 30 day reference period may lead to overestimation of the unemployed and underestimation of the inactive compared with this standard method.

Those *not in the labor force* includes those who are not currently in employment, and are not looking for employment. It also includes those who say that they would like to work, but have not attempted to find work in the last 30 days. Students, those unable to work because of health reasons, those in working age who are retired and not working, and housewives looking after children or other members of the household are classified as not in the labor force. However, if any of the latter categories have part-time work, they are included in the employed category.

ECSTAT categorizes the 'socio-economic status' of the individual and is derived from the variables LFSTAT and A1J112. It has the categories:

---

[35] Note that as the constructed adult data file contains records for all adults, those individuals who were not of working age record a missing value ('.') for LFSTAT.

[36] These inconsistencies have apparently arisen because question A1J112 was designed to indicate the respondent's *perception* of his or her situation, rather than an objective measure, and thus the variable may be more suited for use in sociological rather than economic analyses.

1) Employed - entrepreneur, independent labor activity
2) Employed - farmer, worker
3) Employed - other
4) Unemployed
5) Not in the labor force - pensioner
6) Not in the labor force - student
7) Not in the labor force - disabled
8) Not in the labor force - other

The variable **OCCUP** describes the occupation of those employed in an enterprise, organization, collective or state farm, or cooperative and is derived from the answers to A1J2-A1J5. The variable is coded according to the ILO occupational codes (with some changes to appropriately reflect the occupational system in the Kyrgyz Republic). The codebook for OCCUP is available on request (see appendix C). The variable **OCCUPC** gives the main occupation group and has the categories:

1) Armed forces (OCCUP<1100)
2) Legislators, senior officials and managers (1100≤OCCUP≤1399)
3) Other (1500≤OCCUP≤1599)
4) Professionals (2000≤OCCUP≤2999)
5) Technicians and associate professionals (3000≤OCCUP≤3999)
6) Clerks (4000≤OCCUP≤4999)
7) Service workers and shop and market sales workers (5000≤OCCUP≤5999)
8) Skilled agricultural and fishery workers (6000≤OCCUP≤6999)
9) Craft and related trade workers (7000≤OCCUP≤7999)
10) Plant and machine operators and assemblers (8000≤OCCUP≤8999)
11) Elementary occupations (9000≤OCCUP)

6.3. Household characteristics

The household characteristic variables can be found on the data file CORE. These variables are listed in Table 9 and a brief description is presented below.

HID is the unique household identification number of the household.
MONTH is the month the interview was conducted (AA4_2 in the household questionnaire) and it is equal to 10 (11) if the interview was conducted in October (November).

Table 9: Constructed variables: household characteristics

HID	household identification number
REGION	*oblast* of settlement
SETTLM	type of settlement
SITE	location of settlement
ALTIT	altitude of settlement
LOCAL	locality
N_CHILD	number of children in household
N_ADULT	number of adults
N_CHILD6	number of children aged less than 6 years
N_INVAL	number of invalids
N_PENS	number of pensioners
N_UNEM	number of unemployed persons
N_EMP	number of employed persons
HHSIZE	number of people
MONTH	month of interview

The following four variables were constructed using information on the settlement in which the household was situated (AA2 in the household questionnaire):

- The *oblast* in which the household was situated is recorded by the variable **REGION**, the categories of which are:
  1) Narunskaya *oblast* (AA2 = 1-5)
  2) Talasskaya *oblast* (AA2 = 6-10)
  3) Djalal-Abadckaya *oblast* (AA2 = 11-22)
  4) Issuk-Kulskaya *oblast* (AA2 = 23-31)
  5) Oshckaya *oblast* (AA2 = 32-56)
  6) Chyiskaya *oblast* (AA2 = 57-81)
  7) Bishkek (AA2 = 82-85)
- The variable **SETTLM** records the type of settlement in which the household is situated. It takes the values:[37]
  1) urban settlement
  2) rural settlement
- **SITE** records the geographic location of the household and has the categories:
  1) north (Narunskaya, Talasskaya, Issuk-Kulskaya, Chyiskaya, Bishkek)
  2) south (Djalal-Abadckaya, Oshckaya)
- The altitude of the settlement in which the household is situated is recorded by the variable **ALTIT**.[38]
- **LOCAL** records the 'locality' of the household and has the categories:
  1) Bishkek
  2) other urban
  3) rural

The following variables were created using the household roster data in the household questionnaire and also ECSTAT:

[37] It should be emphasized that this fairly crude dichotomy joins *PGTs* (settlements of the urban type) with settlements which are urban in the Western tradition. PGTs are sometimes smaller than settlements classified as rural, and therefore the term 'urban type' really only refers to the manner in which the settlements are administered, rather than the actual level of urbanity.
[38] The altitude of settlements ranges from 550 to 2453 meters. Altitude may be important in the analysis of poverty, for example, as it reflects the viability of different agricultural activities.

- **N_CHILD** records the number of children (those aged less than 18 years) in the household.
- **N_ADULT** records the number of adults in the household
- **N_CHILD6** is the number of children aged less than six years.
- **N_INVAL** records the number of invalids in the household.
- **N_PENS** is the number of pensioners in the household.
- **N_UNEM** is the number of unemployed persons in the household.
- **N_EMP** is the number of employed persons in the household.
- **HHSIZE** is the number of people in the household.

6.4. Household expenditure variables

Monthly household expenditure variables were constructed using data from the household questionnaire. The monthly household expenditure variables can be found on the data file INCEXP and are listed in Table 10. A description of their construction is presented below.

6.4.1. Adjusting for outlier observations, missing values and coding problems

Outliers were adjusted for at the expenditure sub-aggregate level (defined below). Outliers were defined as those observations which deviated by more than 5 standard deviations from the mean and they were replaced with means using a two-stage procedure. Outlier identification and adjustment was performed within three 'localities' (Bishkek, other urban and rural) and only over the positive observations for each variable. All variables were measured at the household level and in monthly per capita terms.

Table 10: Constructed variables: household expenditure

HID	household identification number
FOODX	food
RENTX	rent
OTHOUSX	other housing expenditure
EDUCULX	education
HEALTHX	medical and health
TRANSX	transport and communications
CLOTHX	clothing
PRITGX	private gifts given
OTHERX	other expenditures
SELFCNX	consumption of home produced agricultural products
TOTHHX	total monthly household expenditure

All missing values were set to zero.

6.4.2. Aggregation of expenditures

If a household was interviewed in November, then expenditures were divided by 1.19 to make them comparable to expenditures made in October.[39] The following 9 expenditure sub-aggregates were constructed (where necessary, with appropriate adjustment to convert expenditures into monthly equivalents):

- Food expenditure (**FOODX**) is the sum of expenditure on: 10 food groups (subsets of questions AE1_1 to AE1_62): breads, potatoes, fruit and vegetables, meats, dairy

[39] This deflator was provided by the survey team.

products, fats, sugar, eggs, fish, and other food; eating out, which is given by question AE9.

- **RENTX** is expenditure on rent (AE18).
- Other housing expenditures (**OTHOUSX**) is the sum of expenditure on: utilities (AE19); heating fuel (AE15_3B, AE15_4B); repair and maintenance of house and contents (AE16_4B to AE16_7B).
- **EDUCULX** is expenditure on education (AE20_1B to AE20_5B, AE20_7B).
- **HEALTHX** is medical and health expenditure (AE15_1B, AE16_9B, AE20_6B).
- Expenditure on transport and communication (**TRANSX**) is the sum of expenditure on: purchase of car, truck etc. (AE14_12C, AE14_13C); transport fuel (AE15_2B); public transport (AE16_1B); taxis (AE16_10B); communication services (AE16_11B).
- Expenditure on clothing (**CLOTHX**) is the sum of expenditure on purchase of clothing (AE11, AE13) and repair of clothing (AE16_2B, AE16_3B, AE16_8B).
- **PRITGX** is the value of private gifts given(AE22_1B to AE22_6B).
- Other expenditure (**OTHERX**) includes: expenditure on alcohol and tobacco (AE1_63C to AE1_68C); expenditure on durables (AE14_1C to AE14_18C, not including AE14_12C and AE14_13C); funeral services (AE16_12B); insurance premiums (AE20_10B) and alimony (AE20_13B).

Total monthly household expenditure (**TOTHHX**) is the sum of the above 9 sub-aggregates (plus consumption of home produced goods,**SELFCNX**, which is equal to SELFCNY described below).

6.5. Household income variables

Monthly household income variables were constructed using data from both the adult and household questionnaires. The monthly household income variables can be found on the data file INCEXP and are listed in Table 11. The construction process involved four stages: calculation of the income of individual household members, calculation of household income derived from the land, calculation of other household income, and calculation of total household income. A description is presented below:[40]

6.5.1. Adjusting for outlier observations, missing values and coding problems

See 6.4.1 above for details of the method adjustment for outliers and missing values.

6.5.2. Income of individual household members

[40] In the construction of the income variables, information from the adult file was cross-checked using summarizing variables in the household data. For example, income from unemployment benefits was calculated as the maximum of the sum of benefits reported in the adult file and total unemployment benefits as reported in the household file. This was done because non-response in individual file could lead to under-estimation of household income.

Data from the adult questionnaire was used to calculate the income received during the last 30 days by the individual members of the household. The following household income variables were calculated by summing over household members:

- Salary income, net of taxes (**WAGEMY**) which is equal to salary on the main job (A1J16) plus salary from any additional jobs (A1J33).
- Income from entrepreneurial activities and individual labor activities (**SLFENAY**) was found by the following process.[41] First the entrepreneurial income from three sources was derived. Production income (PIND) is equal to the value of finished goods produced (A1J41) less production expenses (A1J43). Trade income (TIND) is equal to the value of goods sold (A1J53) less the value of goods purchased (A1J55) and relevant expenses (A1J57). Services income (SIND) is equal to the amount received for services rendered (A1J65) less relevant expenses (A1J67). To calculate the individual's share of income from entrepreneurial activities the sum of the incomes from the three sources was multiplied by the individual's share in the business (A1J70). TOTIND was then calculated by adding to this any income earned from other individual labor activities over the previous month (A1J77).
- Elderly pension benefits (**ELDPENY**) and other pension benefits (**OTHPENY**) were derived from answers to A1J88 and A1J90.
- Unemployment benefits (**UNEMPY**) were set equal to A1J109.

Table 11: Constructed variables: household income

HID	household identification number
WAGEMY	salary income
SLFENAY	income from entrepreneurial and individual labor activities
ELDPENY	elderly pension benefits
OTHPENY	other pension benefits
UNEMPY	unemployment benefits
SELFCNY	consumption of home produced agricultural products
SELFEAY	sales of home produced agricultural products
FAMILYY	childcare benefits
SOCASSY	social assistance benefits
OTHSOCY	other social transfers
PRITRY	private gifts received
OTHERY	other income
TOTHHY	total monthly household income

6.5.3. Income from livestock sales

The calculation of livestock income was fairly crude. Because there is no summary variable for income from sales of livestock, in deriving this income source it was necessary use the livestock prices provided by each household. Questions AD14 to AD140 provide

[41] Negative values of SLFENAY were transformed into zeros during the data cleaning process, because it was felt that since there is a tendency to overstate self-employment expenses, a reported negative self-employment income was probably not an accurate indication of the household's financial situation. This also applied to sales of home produced agricultural products (SELFEAY).

information on seven types of livestock and income was calculated for each type. The calculation of monthly income from sales of sheep, for example, is equal to:

((AD51 - AD59)*AD62)/12

where AD51 is the number of live sheep or lambs sold over the last 12 months, AD59 is the number of live lambs or sheep bought over the last 12 months, and AD62 is the household's estimate of the current price of an adult sheep.

A problem with this method is that assumes that the prices relevant to buying and selling sheep are the same. This is unlikely to be true, but there did not appear to be a better way to estimate income from sales of livestock. Another problem is that no information on expenses associated with rearing livestock were provided (however, this is also true of the calculation of all sources of income derived from the land). The total income from sales of livestock (ILSTOCK) was calculated as the sum of each of the components.

6.5.4. Income from animal and harvest products

The calculation of income from the sale and consumption of home produced animal products used information contained in questions AD141_1 to AD141_7. The survey contains information on seven types of animal products: meat, poultry, milk, eggs, honey, wool, and pelts.

The first step was finding the relevant price for these products. It would have been preferable to use farmgate prices for the calculation of income from animal and harvest products. However, a problem with this was that there was not enough information to derive farmgate prices for the calculation of income from the sale of harvest products (there was no information on the value of sales of harvest goods). In the interest of uniformity, it was therefore decided to use market prices for both animal and harvest products. For most of these goods, the country-wide market price (see appendix G) was used. It would have been preferable to use regional market prices, but these were not available at the time.[42]

The calculation of the value of monthly sales and consumption of home produced animal products is now illustrated for the example of meat. The value of the consumption of meat by the household over the previous month was calculated as:

(PMEAT*AD141_1D)/12

where PMEAT is the price of meat and AD141_1D is the quantity consumed by the household over the past 12 months. The following summary variables were calculated: SCANPROD is the total value of consumption of home produced animal products and IANPROD is the value of total sales of home produced animal products (AD148).

[42] For wool and pelts, no market prices were available. For these products, 'farmgate' prices (i.e. the price actually received by the household) were used. For wool, for example, the farmgate price was calculated by dividing the household's estimated sale value of the produce (AD141_6C), by the quantity produced (AD141_6B).

The calculation of the value of sales and consumption of the 26 harvest products was exactly analogous to that used for animal products. The calculation of the summary variables was also analogous. The following summary variables were produced: the total value of consumption of home produced agricultural products (SCHARV) and the value of total sales of home produced harvest products (IHARV), which is equal to AD13 plus AD151.

The following sub-aggregates summarizing total income derived from the land were calculated. The total value of consumption of home produced agricultural products (**SELFCNY**) is equal to the sum of SCHARV and SCANPROD. The value of total income from sales of products and animals raised on the land (**SELFEAY**) is the sum of ILSTOCK, IHARV, and IANPROD.

6.5.5. Other household income

Other household income was summarized as follows. Note that if households received 'in-kind' payments then they were asked to estimate the value in soms:

- **FAMILYY** is the sum of four types of childbirth and childcare benefits (AF9_1B through AF9_4B);
- Social assistance, or subsidies from local authorities (**SOCASSY**) is AF6.
- Other social transfers (**OTHSOCY**), which includes stipends (AF14_3B) and sickness benefits (AF14_4B).
- Private transfers (**PRITRY**), AF13_1B to AF13_7B.
- Other income (**OTHERY**), which includes investment income (AF14_9B, AF14_10B), income from private organizations (AF13_8B to AF13_10B) and alimony payments (AF14_12B).

6.5.6. Total household income

Total monthly household income, TOTHHY, was calculated as the sum of the above 12 income sub-aggregates.

6.6. Poverty line and poverty indicators

The poverty line used for analysis of the 1993 KMPS was calculated on caloric needs and customized for each household to reflect its demographic composition.[43] Table 12 below lists the poverty line and poverty indicator variables which are available on the constructed data file KYGPOV. This sub-section briefly outlines the construction of the variables on these files (for a more detailed discussion of the construction of the food baskets and poverty line, see appendix E).

[43] See Popkin (1994b) for more details on its construction. It should be noted that in constructing the poverty line there was no allowance for the potential presence of household economies of scale.

Table 12: Constructed variables: poverty line and poverty indicators

HID	household identification number
LCSBLNE	low-cost poverty line
HCSBLNE	high-cost poverty line
RLCSBLNE	regional specific low-cost poverty line
RHCSBLNE	regional specific high-cost poverty line
POOR3E	poverty indicator (low-cost poverty line as reference)
POOR4E	poverty indicator (high-cost poverty line as reference)
VPOOR3E	severe poverty indicator (low-cost poverty line as reference)
VPOOR4E	severe poverty indicator (high-cost poverty line as reference)
RPOOR3E	poverty indicator (region specific low-cost poverty line as reference)
RPOOR4E	poverty indicator (region specific high-cost poverty line as reference)
RVPOOR3E	severe poverty indicator (region specific low-cost poverty line as reference)
RVPOOR4E	severe poverty indicator (region specific high-cost poverty line as reference)

6.6.1. Poverty line

There are four poverty line variables on the constructed data file. The variables **LCSBLNE**, and **HCSBLNE** are poverty lines derived from the low-cost and high-cost food baskets respectively, where the costs of the baskets were calculated using Kyrgyz Republic-wide market prices. For each household, the poverty line is calculated by summing the subsistence income levels for the individual members of the family (found in Table 14 in appendix E), where these income levels are derived based on the assumption that 79.2 percent of expenditures were devoted to food. The variables **RLCSBLNE**, and **RHCSBLNE** are poverty lines derived from the low-cost and high-cost food baskets respectively (found in tables 15 and 16 in appendix E), where the costs of the baskets were calculated using *oblast*-specific market prices.[44] While it would be preferable to have rural and urban-specific poverty lines, the necessary price data was not available.

6.6.2. Poverty indicators

The remaining eight variables on the constructed data files (apart from HID) are all dummy variables, where a value of one indicates that the household is poor, based on a particular poverty line and poverty criterion. Compared with household income, the household expenditure variable is likely to be a more accurate indicator of the standard of living of a particular household, and for this reason the majority of the poverty analysis in involved comparing TOTHHX with various versions of the poverty line.

The variable **POOR3E** (**POOR4E**) takes a value of one if TOTHHX is less than LCSBLNE (HCSBLNE). The variable **RPOOR3E** (**RPOOR4E**) takes a value of one if TOTHHX is less than RLCSBLNE (RHCSBLNE). To help identify those households facing severe poverty, the variable **VPOOR3E** (**VPOOR4E**) takes a value of one if TOTHHX is less than half of LCSBLNE (HCSBLNE). Similarly, the variable **RVPOOR3E** (**RVPOOR4E**) takes a value of one if TOTHHX is less than half of RLCSBLNE (RHCSBLNE).

[44] For details on the prices used, see appendix G.

Appendix A. How to Obtain the KMPS Data

The KMPS data are the sole the property of the Government of the Kyrgyz Republic. The World Bank will be provided with unrestricted use of the KMPS data and is authorized by the Government of the Kyrgyz Republic to transfer the KMPS data to third parties. The data from the KMPS will be made available for use by the general research community subject only to the restrictions that:

i) data users shall not transfer them to third parties;
ii) in all uses of the data, due recognition of their source shall be made;
iii) the researcher shall make copies of all publications stemming from the data available to NATSKOMSTAT

NATSKOMSTAT
374 Frunze Street
Bishkek 720000
Kyrgyz Republic

Requests for the KMPS data should include a detailed description of the intended research and may be directed to:

Living Standards Measurement Study
Poverty and Human Resources Division
Policy Research Department
The World Bank
1818 H Street, N.W.
Washington, DC 20433
USA
email: LSMS@worldbank.org

There is a nominal fee associated with the data sets. The World Bank provides them on the 3 1/2" diskette, in SAS portable (XPORT engine), STATA (version 2.1), or ASCII files. The Poverty and Human Resources Division of the World Bank 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 or used for other research. Other researchers must contact the World Bank directly for access to these data. Any infringement on this policy will result in the denial of future access to World Bank LSMS data.

Appendix B. List of KMPS Data Sets Available

The following data sets are available on 3 1/2" diskette. All are available in SAS portable (XPORT engine), STATA (version 2.1) and ASCII files.

Data set name	Description	Variables in data set	Number of variables	Number of observations
Raw data				
KHHLD	Household Questionnaire (excluding module B)	COMMID, HID, AA1--AA7, AC1--AG6	901	1937
KINDIVH	household roster (module B of Household Questionnaire)	HID, AB1_2--AB169_15	234	1937
KADULT	Questionnaire for Adults (excluding module P)	HID, PID, A1H2--A1O46_2, A1Q1_1--A1R6	434	5647
KADIET	adult nutrition data (module P of Questionnaire for Adults)	HID, PID, A1P1_X_6, A1P1_X_7[a]	23	18284
KCHILD	Questionnaire for Children (excluding module P)	HID, PID, A1H2--A1O44_2, A1Q1_1--A1R6	246	3421
KCHDIET	child nutrition data (module P of Questionnaire for Children)	HID, PID, A1P1_X_6, A1P1_X_7[a]	27	12860
KPRICE1	Survey of Availability and Prices of Food Products and Fuel	COMMID, AYCOV2--AYD5C3	568	212
KPRICE2	Survey of Availability and Prices of Food Products and Fuel (continued)	COMMID, AYE1ID--AYE5_26E	886	212
KPRICE3	Survey of Availability and Prices of Food Products and Fuel (continued)	COMMID, AYE6ID--AYEE_3E	1000	212
KCOMM	Survey of Community and Social Infrastructure	COMMID, A1XC2--A1X90_7	273	213
Constructed data				
KINDIV	household demographic and relationship variables (individual level)	See Table 7	24	9547
CONADULT	adult characteristics	See Table 8	11	5647
CORE	household characteristics	See Table 9	15	1937
INCEXP	income and expenditure	See Tables 10 and 11	25	1937
KYGPOV	poverty line and poverty indicators	See Table 12	13	1937

Notes: KPRICE1, KPRICE2, and KPRICE3 can be merged using COMMID.

[a] - where 'X' is a number ranging from 1-10 (only for KADIET and KCHDIET).

Appendix C. List of Related Documents

The following documents can be obtained from the World Bank at a cost of 5 cents per page for photocopying.

A. Questionnaires (free of charge)

1. Household Survey
2. Individual Survey: Adult
3. Individual Survey: Child
4. Price and Food Availability Survey
5. Community and Social Infrastructure Survey

B. Interviewer Instruction Manuals

1. Instructions for the Household Survey, 38 pages
2. Instructions for the Price Survey, 11 pages

C. Other

1. 'Overall Procedures,' (technical note by survey team), 12 pages
2. 'Sample of Kyrgyzstan,' (technical note by survey team), 42 pages
3. Classification of Occupations, 1994 (based on ILO's ISCO-88), 8 pages
4. Cyrillic codes (questions A1J36C, A1J46C and A1J60C), 18 pages

Appendix D. List of Papers and Reports

The following is a list of papers, published and unpublished, and research in progress that make use of the KMPS data insofar as we are aware as of June 1995. This list is provided to guide researchers who are seeking to build on, but not reproduce work that has already been done. Copies of all papers and publications resulting from the analysis of KMPS data sets should be sent to the World Bank at the address listed in appendix A.

Ackland, R. (1995), 'A Preliminary Analysis of Labor Supply in the Kyrgyz Republic,' mimeo. Yale University.

----------and Falkingham (1996), 'Economic Transition and the profile of poverty in Kyrgyzstan,' in J. Falkingam et al. (eds.) *Household Welfare in Central Asia* , London: Macmillan (forthcoming).

Cox, D., Jimmenez, E. and Jordan, J (1994), 'Family Safety Nets and Economic Transition: A Study of Private Transfers in Kyrgyzstan,' mimeo. The World Bank.

Evandrou, M. and Falkingham, J. (1994), 'Social Service Utilisation in Kyrgyzstan: A Preliminary Analysis from The Multipurpose Poverty Survey,' mimeo. The World Bank.

Falkingham, J. (1994), 'Social Assistance and the Social Safety Net in Kyrgyzstan,' mimeo. The World Bank.

Marnie, S. (1994), 'The Kyrgyz Labor Market in Autumn 1993: The Results of the Kyrgyz Multipurpose Poverty Survey,' mimeo. The World Bank.

Popkin, B. (1994a), 'Overview: The Multipurpose Poverty Survey,' mimeo. The World Bank.

---------- (1994b), 'A Subsistence Income Level for Kyrgyzstan,' mimeo. The World Bank.

---------- and Martinchik, A. N. (1994), 'Nutritional Conditions of the Kyrgyz Population, 1993,' mimeo. The World Bank.

Appendix E. The Poverty Line

This appendix provides information on the definition and calculation of the poverty line. More detailed information can be found in the relevant background papers referenced in appendix C.

There are two main approaches to constructing a poverty line. An absolute definition of poverty assumes that it is possible to define a minimum standard of living based on a person's physiological needs for food, water, clothing and shelter. The relative approach, however, defines poverty in relation to a generally accepted standard of living in a specific society at a specific time and goes beyond basic physiological needs. The absolute and relative approaches to defining poverty have advantages and disadvantages. An advantage with the absolute approach is that there are reasonably objective norms, while with the relative approach the decisions concerning what is an acceptable minimum become much more subjective. The choice between the two approaches largely rests upon the context and purpose to which the poverty line will be put. In the Kyrgyz Republic today, the priority is to identify those who are most in need rather than those who are disadvantaged relative to others but who can still afford the basic necessities. For this reason, an absolute approach to calculating the poverty line was adopted.

The method used to define the poverty line followed the common approach of estimating the cost of a basket of goods which is chosen to ensure that basic consumption needs are met. Individual dietary intake data collected from the KMPS were used to find a minimum cost food basket which, in addition to achieving a required calorific level, contained a mixture of foods which adequately reflected the Kyrgyz diet. Two food baskets, the low-cost and high-cost food baskets, were developed. The former is a food basket which reflects a more austere diet and which deviates more from the current consumption pattern of low-income Kyrgyz than does the high-cost one. However, it should be noted that both food baskets provide the same level of nutrients and allow for adequate growth and activity.

The KMPS data indicated that for low income Kyrgyz households, 79.2 percent of expenditures were devoted to food. Non-food consumption was allowed for by dividing the cost of the food basket by this share to estimate a subsistence income level associated with the food basket. This subsistence income is the poverty line which was subsequently used in the poverty profile to determine whether a particular household was poor or not. The poverty line calculated thus specifically reflects the demographic composition of the household. Table 14 below presents costs of the food baskets and the corresponding subsistence income levels for different demographic groups. The poverty line for a particular household is the sum of the poverty lines for the individual members. Tables 15 and 16 present the high- and low-cost food baskets by *oblast* (using *oblast* level food price data described in appendix G). It would have been preferable to have rural- and urban-specific poverty lines, however the necessary price data was not available.

Derivation of the minimum cost food basket

The food basket was chosen so as to minimize total cost, subject to dietary and other constraints. An analysis of the nutrition data in the KMPS revealed that 42 food items account for 98 percent of total energy in the average Kyrgyz diet. These foods were selected for inclusion in the food basket.[45]

[45] For more details on the construction of the food basket, see Popkin (1994b).

The following linear program was solved:

$$\text{Minimize } C = \sum_{j=1}^{52} c_j x_j$$

subject to:

$$\sum_{j=1}^{8} a_{ij} x_j \leq r_i, \quad i=1,...,8$$

and:

$$x_j \geq 0, \qquad j=1,...52$$

Where:

C is the total cost of the food basket,
c_j is the price of the jth food,
x_j is the amount of the jth food in the food basket (the x's are the choice variables in this problem),
r_i is a parameter in the ith dietary constraint,
a_{ij} is the coefficient on the jth food in the ith dietary constraint.

The dietary constraints are :

1. *daily energy intake constraint* - the coefficients were the amount of energy in kcal in each food item.
2. *daily protein intake constraint* - the coefficients were the total amount of protein in grams in each food item.
3. *daily animal protein intake constraint* - the coefficients were the amount of animal protein in each food item.
4. *daily plant protein intake constraint* - the coefficients were the amount of plant protein in each food item.[46]
5. *daily animal fat intake constraint* - the coefficients were the amount of animal fat in each food item.
6. *daily plant fat intake constraint* - the coefficients were the amount of plant fat in each food item.
7. *daily carbohydrate intake constraint* - the coefficients were the amount of carbohydrate in each food item.

[46] This sub-division for proteins was made so as to achieve balance in the consumption of both animal and food protein. The same applies to the sub-division between animal and plant fat.

Table 13: Annual food basket and subsistence income level, October 1993

	Low cost food basket	High cost food basket	Low cost subsistence income	High cost subsistence income
Average person	522	830	659	1048
1-3 yrs	409	468	516	591
4-6 yrs	442	678	558	856
7-13 yrs	534	841	674	1062
14-17 yrs	539	957	681	1208
Male, 18-59 yrs	593	1017	749	1284
Female, 18-54 yrs	523	814	660	1028
Pensioner	484	779	611	984

Source: Popkin (1994b). Soms per year.

Table 14: High-cost food basket and subsistence income level, by oblast, October 1993

	Narunskaya		Talasskaya		Djalal-Abad'		Issuk-Kul'		Oshckaya		Chyiskaya		Bishkek	
	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.
1-6 yrs	455	575	496	626	449	567	389	491	448	565	465	587	520	657
7-10 yrs	675	852	750	947	652	823	564	712	658	830	684	864	749	946
11-13 yrs	826	1042	945	1193	814	1028	708	893	827	1044	850	1073	926	1169
14-17 yrs	939	1185	1075	1357	926	1169	805	1016	941	1188	966	1220	1053	1329
Male, 18-59 yrs	992	1253	1143	1443	1002	1265	867	1095	1009	1274	1023	1291	1111	1403
Female, 18-54 yrs	795	1003	915	1156	802	1013	695	877	808	1020	819	1034	890	1123
Pensioner	780	985	868	1096	761	960	692	874	760	960	801	1012	844	1066

Source: Survey team. The subsistence income level is calculated as the value of the food basket divided by 0.792 (under the assumption that 79.2 percent of expenditures were devoted to food). Soms per year.

Table 15: Low-cost food basket and subsistence income level, by oblast, October 1993

	Narunskaya		Talasskaya		Djalal-Abad'		Issuk-Kul'		Oshckaya		Chyiskaya		Bishkek	
	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.	basket	inc.
1-6 yrs	401	507	430	543	390	493	340	429	384	485	408	515	447	564
7-10 yrs	438	553	466	589	410	518	364	460	408	515	441	557	486	614
11-13 yrs	545	688	569	719	510	644	440	556	508	641	546	690	590	745
14-17 yrs	543	686	593	749	508	642	458	578	512	647	549	693	603	761
Male, 18-59 yrs	603	761	677	854	557	704	511	645	564	712	601	758	665	839
Female, 18-54 yrs	534	674	588	742	497	627	445	561	500	632	532	672	589	744
Pensioner	493	623	542	684	454	574	415	525	457	577	491	620	547	691

Source: Survey team. The subsistence income level is calculated as the value of the food basket divided by 0.792. Soms per year.

Appendix F. Codes Not in the Questionnaires

The majority of the codes for categorical variables are either in the questionnaires or else they are in the interviewer instruction manuals. For those variables with codes which are not in the above sources, the codes are provided below:

Adult and Child Questionnaires

The variable A1I5 in the adult and child questionnaires has the following codes:

1 RSFSR	14 Turkmenia	27 Ivanovsk *oblast* Russia	40 North Korea
2 Ukraine	15 Mari ssr	28 Primore Russia	41 Krasnoyarsk krai
3 Byelorussia	16 Tatar assr	29 Tiumenskaya *oblast*	42 Samarskaya Guberniya
4 Moldavia	17 China	30 Poland	43 Moscow
5 Lithuania	18 Kemerovsky *oblast*	31 Amurskaya *oblast*	44 Karachaevo-chrkess ao
6 Latvia	19 Bashkir assr	32 Perm *oblast*	45 Irkutskaya *oblast*
7 Estonia	20 Voronezhskaya	33 Udmurt Republic	46 Sverdlovsk *oblast*
8 Georgia	21 Crimea	34 Komi assr	47 Dzumgalsk
9 Armenia	22 Orenburg *oblast*	35 Khakassiya-	48 KNR
10 Azerbaijan	23 Novosibirsky *oblast*	36 Dzhambul *oblast*	49 Voronezhskaya *oblast*
11 Tadzhikistan	24 City Cheboksary	37 Chechnia-Ingushskaya	
12 Uzbekistan	25 Altai Krai	38 Mordovskaya assr	
13 Kazakhstan	26 Krasnodar Krai	39 Yakutia	

The variable A1I15 in the adult and child questionnaires has the following codes:

1 Kygyz	9 Korean	17 Polish	25 Moldavian	33 Mixed (Pidgin?)
2 Russian	10 Tadzhik	18 German	26 Deaf mute	34 Kumyk
3 Uzbek	11 Tatar	19 Dungan	27 Mordvin	35 Farsi
4 Kranian	12 Turkish	20 Arab	28 Udmurt	36 Karachaevetz
5 Turkmen	13 Does not talk (child)	21 Kalmyk	29 Lezgin	37 Donskoi Kazakh
6 Kazakh	14 Uygur	22 Bashkir	30 Kurd	38 Balkarsky
7 Belorussian	15 Dungan	23 Bolgar	31 Laketz	39 Lithuanian
8 Chinese	16 Afgan	24 Azerbaidzhan	32 Jewish	

The variables A1J51_1 to A1J51_5 in the adult questionnaire have the following codes:

1 China
2 CIS (SNG)
3 Uzbekistan
4 Ukraine
5 India
6 Russia
7 Does not know
8 Refused
10 Turkey
11 Poland
12 Pakistan

The codes to variablesA1J36C, A1J46C and A1J60C in the adult questionnaire are in Cyrillic and are available as an attachment (see Appendix C).

The variables A1J2-A1J5 describe the occupation of adults. Information from these questions was used to code the variable OCCUP, which is available on the constructed adult file.

Household questionnaire

Marital status is described by ABX_7 (where X ranges from 1 to 16). For the categories of this variable, see the constructed variable MARITAL, described in sub-section 6.1.

Appendix G. Food Price Data

Food price data were used in the construction of both the poverty line and also the household income and expenditure variables. This appendix presents the food price data and describes their collection.

Kyrgyz Republic-wide food prices

The construction of the Kyrgyz Republic-wide poverty line required national market prices for each item in the food basket. These prices were also used in calculating the value of home produced agricultural products sold and consumed by the household. The prices used were those collected by GOSKOMSTAT for October, 1993 and are listed in Table 17 below.

Oblast food prices

Poverty lines were also constructed for each *oblast*, to reflect regional differences in food prices. While the KMPS collected food price data at the community level, this data was not used to find the *oblast*-level food price data.

Table 18 below shows the food price data used for the costing of the food baskets used for constructing the regional poverty lines. It was not possible to get the relevant food prices for every *oblast* and the following describes the imputations which were used to 'fill in the gaps'.[47]

In general, the imputations involved the following formula:

$$P_{ik} = (P_{il}/P_{jl})*P_{jk}$$

where i is the *oblast* for which there is not price data for food item k, j is a 'similar' *oblast* (i.e. both *oblasts* are in the south or north), and l is a food which is a close substitute for food k and for which there is price information for both *oblast* i and *oblast* j.

The following details the price imputation for specific foods:

1) Prices for lepeshka (LEP_1S), high quality bread (BREAD_HS), and bread (BREAD_BL) were missing for all *oblasts* except Bishkek. These prices were imputed using the above formula and the price distribution for first grade bread (BREAD_1S).

2) Prices for second grade wheat flour (WHFLO_2) were taken from the price list for Kyrgyzstan as they were missing for all *oblasts*.

3) Prices for first grade wheat flour (WHEATFLO) were missing for 4 *oblast*. Under the assumption of approximate equality of prices in the south, the imputed price for Oshckaya *oblast* was that recorded for Djalal-Abad' *oblast*. Missing prices for other *oblasts* were imputed using the above formula and the price distribution for bread, where Chyiskaya was used as the reference *oblast* (i.e. *oblast* k in the above formula).

4) Prices for high grade wheat flour were imputed based on the bread ratio.

[47] These notes are based on technical notes provided by the survey team. In some parts they are incomplete.

5) The missing price for rice in Oshckaya was imputed using the price of rice in the other southern *oblast*, Djalal-Abad'. The price of rice in Issuk-Kulskaya was set at the average price in the other northern *oblasts*. The same calculation provided the price of other grains in Narunskaya.

6) As the price of beans was missing for all *oblasts*, the Kyrgyz Republic-wide price was used for each *oblast*.

7) The missing prices for other vegetables were imputed by the above formula, using the price distribution of roots and Bishkek as the reference *oblast*.

8) The prices for dry fruit were imputed based on the price ratio for fresh fruit.

9) The price for sugar in Issuk-Kulskaya is the average of sugar prices in the northern *oblasts*.

10) Prices for cakes were imputed using the price ratio for sugar.

11) Prices for various types of meat were imputed using the price ratio for MUTTON.

12) The price of fish was missing for all *oblast* except Bishkek. The prices were imputed based on the MEAT price ratio.

13) Prices for cream were imputed using the price ratio for milk where Bishkek was the reference *oblast*.

14) Prices for cheese were set at the average of the prices for Bishkek and Issuk-Kulskaya.

15) The prices for vegetable oil and margarine were taken based on the ratio of prices for BUTTER.

16) Prices for eggs were recalculated, as in the food basket we use grams per day but the price for eggs we got in som per ten. Therefore, knowing that the weight of the egg is 40 grams, EGGS prices were recalculated.

<u>Other food prices used</u>

In the construction of the income and expenditure variables, a deflator was used. This was provided by the survey team.

Table 16: Kyrgyz Republic-wide market food prices, October 1993 (soms per kg)

Bread Products	
total flour	na
rye	na
wheat (1 sort)	0.56
flour wheat (highest quality)	0.74
pasta	
pasta (ordinary)	2.50
pasta (highest quality flour)	na?
'straw' pasta (high quality flour)	1.50
vermicelli	1.71
vermicelli (highest quality wheat)	1.48
rice	1.51
other grains (not rice)	
semolina	1.10
millet	0.76
buckwheat	5.00
oatmeal	0.93
pearl-barley	0.60
maize	0.84
beans	na
Potatoes	0.81
Vegetables (and Melons)	
cabbage	1.10
cucumber	1.75
tomatoes	1.12
edible roots	
onion	1.10
garlic	2.83
beet-root	0.75
carrot	0.86
pumpkin	0.50
aubergine	0.81
melons	
water-melon	0.58
melon	1.20
Fruits and Berries	
apple	2.06
pear	2.06
plum	2.40
grapes	2.00
orange	na
mandarin	na
lemon	na
Eggs(10)	2.97
Vegetable Oil and Margarine	
cotton-seed oil	10.66
margarine	8.95

Sugar, Pastries and Confectionery	
sugar	4.68
pastries	
honey	13.70
ordinary pastry (baking)	6.21
dry baking	4.55
cakes	5.54
wafer	8.41
splice-cake	3.32
sponge-cake	8.35
short bread (cake)	5.90
caramel	7.17
toffee	8.80
milk sweet	10.00
Meat and Meat Products	
beef	5.28
lamb	5.98
pork	7.00
poultry	
chicken	10.00
duck	5.20
chicken (other type)	11.00
sausage and smoked meat	
boiled sausage (1 sort)	7.00
boiled sausage (2 sort)	6.24
frankfurter (1 sort)	10.60
small sausage	8.30
liver sausage (3 sort)	2.34
sub-products	
pork-liver	2.50
cow and pig feet	2.00
Fish and Fish Products	
live fish	5.60
tinned fish	
natural with oil	2.80
sprats in oil	3.42
fish paste	2.00
Milk and Milk Products	
pasteurized milk	
3.2-3.5 % fat	0.59
1 % fat	0.46
sour cream	3.62
cream	2.10
butter	11.63
cottage cheese products	
fat	1.82
non-fat	1.19
cheese	
solid cheese	10.25
processed cheese	5.22
processed cheese paste	5.60

Source: Survey team

Table 17: Prices for food items by oblast, October 1993

	Narunskaya (N)		Talasskaya (N)		Djalal-Abad' (S)		Issuk-Kul' (N)		Oshckaya (S)		Chyiskaya (N)		Bishkek (N)	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B
LEP_1S		0.70		0.70		0.53		0.55		0.53		0.56	0.88	0.88
BREAD_1S	0.64	0.64	0.64	0.64	0.48	0.48	0.50	0.50	0.48	0.48	0.51	0.51	0.80	0.80
BREAD_HS		0.74		0.74		0.56		0.58		0.56		0.59	0.93	0.93
BREAD_2S		0.42	0.48	0.48		0.32		0.33		0.32		0.34	0.53	0.53
BREAD_BL		0.72		0.72		0.54		0.57		0.54		0.58	0.6	0.60
WHFLO_2		0.56		0.56		0.56		0.56		0.56		0.56		0.56
WHEATFLO		0.77		0.87	0.54	0.54	0.54	0.54		0.54	0.61	0.61		0.96
WFLOTHIN		0.80		0.80		0.60		0.63		0.60		0.64		0.84
PASTA	1.40	1.40	2.16	2.16	1.68	1.68	1.70	1.70	1.88	1.88	1.29	1.29	1.60	1.60
RICE	1.22	1.22	2.00	2.00	1.50	1.50		1.63		1.50	1.70	1.70	1.60	1.60
OTHERGRA		0.90	0.43	0.43	1.03	1.03	0.57	0.57	0.60	0.60	0.63	0.63	1.96	1.96
BEANS		2.15		2.15		2.15		2.15		2.15		2.15		2.15
PATATOES	0.65	0.65		1.10	1.08	1.08	0.49	0.49	0.60	0.60	1.02	1.02	0.94	0.94
CABBAGE	1.30	1.30	1.00	1.00	1.63	1.63	0.78	0.78	1.13	1.13	1.15	1.15	1.02	1.02
CUCTOM	1.50	1.50	2.13	2.13	0.95	0.95	1.69	1.69	0.90	0.90	1.60	1.60	1.30	1.30
ROOTS	1.75	1.75	2.13	2.13	1.09	1.09	1.33	1.33	1.99	1.99	1.62	1.62	1.57	1.57
OTHERVEG		1.18		1.44	1.00	1.00	2.41	2.41	1.77	1.77	1.50	1.50	1.06	1.06
FRESH FRUIT	2.30	2.30	3.00	3.00	1.50	1.50	0.80	0.80	1.88	1.88	2.00	2.00	2.95	2.95
DRY FRUIT		16.39		21.38	10.00	10.00		5.70	4.38	4.38	14.25	14.25	10.67	10.67
SUGAR	5.40	5.40	6.00	6.00	3.50	3.50		5.23	5.00	5.00	5.00	5.00	4.50	4.50
CAKES		7.80		8.67	9.60	9.60		7.55		9.60		7.22	6.50	6.50
CONFECT	6.00	6.00	5.57	5.57	5.22	5.22	4.80	4.80	4.14	4.14	6.79	6.79	6.51	6.51
BEEF		7.24	6.50	6.50	5.00	5.00	3.50	3.50	5.25	5.25	5.00	5.00	5.57	5.57
MUTTON	4.50	4.50	5.50	5.50	7.00	7.00	4.25	4.25	7.00	7.00	5.93	5.93	6.68	6.68
PORK		6.99	8.00	8.00	7.50	7.50	5.13	5.13	7.00	7.00	7.53	7.53	7.29	7.29
POULTRY		15.82		15.82	16.50	16.50	20.00	20.00		16.50	14.65	14.65	12.81	12.81
SAUSAGES		7.68		7.68	5.90	5.90	7.17	7.17	7.27	7.27	7.25	7.25	8.63	8.63
SUBPROD		2.32		2.66		2.49		1.70		2.32	2.50	2.50		2.42
FISH		7.83		7.83		6.02		7.31		7.41		7.39	8.80	8.80
MILKWHOL	0.61	0.61	0.56	0.56	0.60	0.60	0.46	0.46	0.44	0.44	0.67	0.67	0.67	0.67
CREAM		1.91		1.76		1.88		1.44		1.38		2.10	2.10	2.10
BUTTER	10.46	10.46	11.07	11.07	14.90	14.90	12.49	12.49	11.00	11.00	11.00	11.00	10.50	10.50
COTTAGE	2.11	2.11	1.80	1.80	1.75	1.75	1.54	1.54	1.20	1.20	2.54	2.54	1.50	1.50
CHEESE		10.25		10.25		10.25	10.20	10.20		10.25		10.25	10.30	10.30
EGGS	2.70	6.75	4.75	11.87	3.60	9.00	1.90	4.75	3.70	9.25	2.50	6.25	2.80	7.00
OILSMARG	8.40	8.40		11.12	9.30	9.30	11.50	11.50		9.30		11.12	13.45	13.45
MARGARINE		9.55		16.79		8.00		6.72	8.00	8.00		8.84	9.90	9.90
SEEDNUT		10.64		10.64		10.64		10.64		10.64		10.64		10.64
TOM PUREED		12.00		12.00		12.00		12.00		12.00		12.00		12.00
VODKA		10.18		10.18		10.18		10.18		10.18		10.18		10.18
TEA	16.00	16.00	35.00	35.00	16.25	16.25		20.55		16.25	20.00	20.00	11.20	11.20

Source: Survey Team.

Notes: The letters in parentheses after the *oblast* name denotes whether the region is in the north or the south. Column A contains the original price data, while column B contains the price data including imputations.