SAMPLE OF KYRGYZSTAN
WORLD BANK HOUSEHOLD SURVEY, 1993

The purpose of this sample was to represent all households and individuals in Kyrgyzstan in Fall, 1993. In drawing the [this] sample, we strove to approximate a probability sample--that is, a sample in which every household had some non-zero calculable random chance of falling into the sample. This, for example, eliminated the possibility of using quotas or of choosing "typical" cities or regions. Both of these approaches generate inferior samples to which the findings of theoretical statistics cannot be validly applied. Probability samples, on the other hand, do permit the application of theoretical statistics in estimating how well the sample represents the population. [Conversely, probability samples permit the application...]

In this study, "household" was defined as a group of people who live together in a given domicile, who keep house together, and who share common income and expenditures.[1] Judging from the 1989 census, there were about 856,000 families containing 4,258,000 individuals living in Kyrgyzstan at that time--an average of about five members per family. Though our definition of "household" and the census definition of "family" differ somewhat, the figure 856,000 can be taken as an estimate of the number of households from which our sample was to be drawn.[2] Since our target household sample size was 2,000, to allow for a modest non-response rate of about five percent, we drew a sample of 2,100 households. Incidently, this estimate of non-response, though low by Western experience, proved very accurate. The actual number of completed household interviews was 1,960--a response rate of 93.3%.

As is common in drawing national samples of households, we used a multi-stage procedure. The number of stages depended on whether we were sampling from cities or rural areas (see below). Roughly speaking, at the first stage several of the administrative-territorial jurisdictions of Kyrgyzstan were selected as primary sampling units (PSUs). At the second stage, several of microcensus enumeration districts within the selected PSUs were selected randomly. These secondary sampling units (SSUs) were made available to us through the cooperation of GOSKOMSTAT of Kyrgyzstan in connection with its preparation for a microcensus scheduled for early 1994. In rural areas, villages were used in place of microcensus enumeration districts. Finally, in the third stage, a certain number of households were drawn randomly within the selected SSUs. Interviewers were given no discretion in the selection of households; furthermore, no substitution or replacement was permitted.

---

[1]Children of household members who were under eighteen years of age and who lived elsewhere as students were included as household members. For more detail, consult *Project Kyrgyzstan. Sample Survey of Households. Instructions to the Interviewer* (Bishkek, 1993), pp. 8 and 9. Also consult instructions in the household questionnaires.

[2]The fact that the number of households may well have increased since 1989 has little practical effect under the assumption that the growth was roughly consistent across regions.

In addition to administering a questionnaire about the household to an authoritative member of each of the households drawn in this manner, we also administered separate individual questionnaires to all members of the selected households. (Adults answered questions pertaining to children thirteen years of age and younger.) Since questionnaires were obtained for all members of the households (with some non-response, of course), our procedures in principle yielded a probability sample of *individuals* in Kyrgyzstan, not just a probability sample of *households*.

The remainder of this report provides documentation on precisely how the three stages of sampling were actually carried out.

THE SELECTION OF PRIMARY SAMPLING UNITS

Ideally, PSUs are drawn so as to represent fully the variation in regions while greatly reducing the actual territory which interviewers must cover. This reduces the cost of the survey while, hopefully, only modestly reducing the accuracy of the results. On this basis, oblasts were summarily dismissed as potential PSUs because there are only six in Kyrgyzstan. That is, there would have been little point in drawing, say, three out of six oblasts in the primary stage.

These six oblasts are further divided into a total of 57 administrative-territorial units--that is, raions (counties) as well as some important cities falling directly under the jurisdiction of the oblasts rather than of the raions in which they are located.[3] Even 57 is a relatively small number of units from which to draw PSUs. For technical reasons, one normally wants to draw well over 30 PSUs for any sample.[4] In this case, drawing, say, 30 out of the 57 administrative-territorial units would yield less savings in travel costs than we would have liked. However, we had neither the time nor the data with which to construct smaller PSUs by breaking raions up. Thus, the list of these 57 administrative-territorial units served as the basis of selecting the PSUs.

Self-Representing PSUs

---

[3]In the U.S., an analogy would be Bakersfield's answering directly to the State of California rather than to the commissioners of Kern County, in which it is located.

[4]In most key equations, the absolute number of units selected is important, not the proportion of units selected. All else being equal, selecting 30 out of 100 units is better than selecting 15 out of 25 even though the latter percentage (60%) is twice as large as the former (30%). At the risk of overgeneralizing, anything less than 20 to 30 units can scarcely be considered respectable; many well-known studies use 70 to 100. Where there is tremendous heterogeneity (as in Russia) even more PSUs might be warranted. However, the proportion is relevant to some extent in calculating costs. Furthermore, as the absolute number approaches the total, the proportion should figure into the equations as a correction factor.

The fifty-seven administrative-territorial units fall into two categories: self-representing and non-self-representing PSUs. Self-representing units are those which are so populous that at least some inhabitants would be expected to fall into any random sample of a given size. Sometimes, particularly *important* units (e.g. capital cities) are included with certainty in the sample even though their population size is somewhat smaller than the calculated cut-off size for self-representing units.

On this basis, twelve SR administrative-territorial units were selected with certainty:

> the four (4) raions of the capital, Bishkek
> (which is also the administrative center
> of Chuiskaya Oblast);

> the five (5) other oblast administrative centers
> (each consisting of one raion): Dzhalal-
> Abad; Naryn; Talas; Osh; Balykchi
> (formerly Rybache and Issyk-Kul);

> three (3) other major cities (each consisting of
> one raion): Karakol (formerly
> Przheval'sk); Tokmak, and Kara-Balta.

In these 12 PSUs lived 34.2% of the households of Kyrgyzstan. Thus, 718 of the 2,100 households were drawn from these 12 units. Of course, though they were chosen with certainty, they were represented proportionally (see Table 1 on Page 12).

Non-self-representing PSUs

Forty-five administrative-territorial units remained after the selection of the above twelve units. Forty of these were raions; five were cities under the direct jurisdiction of the oblast in which they were located. These five cities (Uzgen, Tash-Kumyr, Kyzyl-Kiya, Kara-Kul', and Mali-Sai) were combined with the raions in which they are geographically situated. This had the effect of increasing the heterogeneity of those raions--a desirable effect from the standpoint of fieldwork. Thus, forty units remained from which the rest of the PSUs were to be selected.

Since 34.2% of households lived within the self-representing raions, 65.8% of the households (i.e., about 590,000 households) were located in these remaining forty units. So, 1,385 of the 2,100 households were allotted to these non-self-representing units.

These 40 units were stratified on three salient bases by experts in GOSKOMSTAT of Kyrgyzstan: geographical conditions (mountain versus valleys); ethnic composition (pure Kyrgyz; mostly Kyrgyz and Uzbek mixed; mostly Kyrgyz and Russian-speaking); and type of production (agricultural versus agricultural-industrial). This process ultimately yielded six strata, which may be roughly described as follows:

I.   mountains, agriculture and animal husbandry, primarily Kyrgyz
     population.

II. mountains, agricultural and animal husbandry, as well as
     nurseries, primarily Kyrgyz population.

III. mountains, agricultural and industrial, Kyrgyz and Uzbek.

IV. valleys, agricultural, Kyrgyz mixed with Russian-speaking
     population.

V.   valleys and mountains, agricultural, Kyrgyz mixed with Uzbek
     population.

VI. valleys, agriculture and industry, Kyrgyz mixed with some Russian-
     speaking.

The populations of strata II and V were about twice as large as those of
strata I, III, IV, and VI. One accepted approach to ensuring that these
strata were proportionally represented was to create two strata out of both II
and V so that there were ultimately a total of eight strata--that is, six plus
the two extra ones thus created.[5] The result of this procedure is
represented in Table 2, in which the raions have been ordered according to
ascending size within the strata to which they were assigned by experts. (The
eight strata numbers have been indicated with Arabic numerals to distinguish
them from the six strata enumerated above with Roman numerals.) The number of
respondents allocated proportionally to each stratum is indicated in
parentheses. The abbreviation after each raion indicates in which of the six
oblasts the raion is located. Other entries in this table are discussed
below.

From each of these eight strata, two raions were chosen systematically
(except in stratum 7, where the two raions were chosen with certainty since
there was a total of only two raions). Though one raion in each stratum would
have sufficed for some purposes, choosing two yielded technical advantages in
assessing the measurability of the sample after the study is concluded; this
also increased the absolute number of PSUs.

The probability of selecting each raion was made proportional to size
(PPS) within its stratum so that ultimately households of all raions had equal
probability of selection. That is, the larger the population of a raion
within a given stratum, the greater the likelihood it would be selected.
Consequently, to conform to the technical standard, once the two raions were
selected, the respondents within each stratum were divided equally between the
two selected raions even though the population of the raions differed. For
example, in stratum 1, the 189 cases were divided equally between AT-
BASHINSKII and DZHETI-OGUZSKII.

---

[5]Other approaches would have been to draw twice as many PSUs from those
two strata or to draw twice as many respondents from the PSUs. These other
approaches would have given rise only to esoteric differences.

THE SELECTION OF SECONDARY SAMPLING UNITS AND OF HOUSEHOLDS

At the second stage, the population points within selected raions were stratified by whether or not they were "city" (*gorodskoi*) or not. We deemed it exceedingly important to stratify on this dimension to insure that both urban and rural areas were proportionally represented in the final sample: 44.5% urban; 55.5% rural.[6] All twelve self-representing units were either cities or were parts of the capital city, Bishkek. That is, they were entirely urban. On the other hand, though most of the non-self-representing units had both urban and rural population points, a few had no urban points at all. Normally, the number of urban population points in a raion was no more than two or three. If there was only one, of course it was selected. If more than one was available, one was selected randomly for each fifteen urban households required in the sample.[7]

Selection of Households

The procedure for selecting households differed for urban and rural areas. In urban areas, we relied on the microcensus enumeration districts which were just being revised GOSKOMSTAT of Kyrgyzstan. Based on experience from the 1989 census, those were each expected to contain about 400 households.[8] We considered it appropriate to choose eight to ten households from a given census enumeration district; we therefore selected enough enumeration districts to yield the desired number of urban households in a given PSU. They were chosen systematically with equal probability. No substitution was permitted.

---

[6] Actually, the term "city" (or "urban") is a somewhat misleading translation of *gorodskoi* in this case, since many population points counted in official statistics as cities or as "settlements of the city type" (PGTs) have populations of only 5,000 to 10,000. Also, sometimes settlements counted as villages can be found that are more populous than other settlements which are treated as being cities. In any case, according to an English understanding of these terms, Kyrgyzstan is roughly 80% rural.

[7] There were eight non-self representing units that had no urban population points whatsoever. This seemed to present a problem. Although they had no urban population points, they were representing strata in which there *were* urban population points. Given the marked importance of this dimension, Mikhail Kosolapov chose to make sure that those urban points in each strata were proportionally represented by taking urban points from non-selected units within the same strata. This was done in a total of 91 of the 2,100 cases. Subsequent discussions have revealed that it was unnecessary to make the adjustment in this way, and that the adjustment will complicate the analysis of the sample *qua* sample (i.e. its measurability). However, these 91 households were also chosen randomly, so we are certain that it did not corrupt the results.

[8] In the 1989 census, there were 11,814 enumeration districts, of which 3,742 were located in so-called urban areas. However, the records from those districts were no longer available.

5

Selection of Households in Villages

Ethnic composition is quite salient in differentiating Kyrgyz villages. We therefore took great pains to insure that the ethnic composition of villages was properly represented in the sample. For each raion, there is a list of villages with data about the ethnic composition of each village, so it was possible to compute the overall composition *of villages* in each of the eight strata. For example, in the first stratum, 94.5% were Kyrgyz; 3% were Russian; 0.8% were Uzbek; 1.7% were other. In the 7th, 54.7% were Kyrgyz, 37.1% were Uzbek; 1.5% were Kurds; 5.2% were other.

In the raions falling into the sample, villages were grouped by nationality (ethnicity): Kyrgyz villages; Kyrgyz-Russian villages, Kyrgyz-Uzbek villages, etc. 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 raion was representing. Enough villages were selected so that no more than 18 to 20 households would be in a selected village. Within a given type of village in a given raion, the villages were listed by size, and villages were selected systematically using an appropriate interval.

For example, suppose a given *stratum* had the following composition of *villages*:

Kyrgyz	55%
Uzbek	30%
Russian	12%
Others	3%

Suppose that a *raion* in that stratum was chosen with the following distribution of *villages*:

40 villages Kyrgyz
50 villages Uzbek
10 villages Mixed Russian-Uzbek-other.

Suppose that 69 households had to be drawn from this raion. We would distribute the households as follows:

Kyrgyz	55%	38 households
Uzbek	30%	21
Russian	12%	8
Other	3%	2 households.

Two Kyrgyz villages of the 40 would be chosen (about 19 households in each of the two). Uzbek households are found both in pure Uzbek villages (about 80% of Uzbek households) and in mixed villages (about 20%). For this reason, we would choose 17 of the needed 21 Uzbek households from a pure Uzbek village. The rest of the Uzbeks, Russians, and others (a total of 14 households) would be selected according to chance from a mixed village. Quotas were not used, so we would not attempt (inappropriately) to guarantee the exact distribution of households by ethnicity at this low level.

In sum, of the 40 Kyrgyz villages, two would be chosen; of the 50 Uzbek villages, one pure Uzbek village would be selected; and out of the 10 mixed villages, one would be selected. The selection of villages within ethnic strata was random. The selection of households within villages was systematic using an appropriate interval. It was drawn from the "book of households" that lists households in rural areas.

CONCLUSION

As is normally the case, the actually number of stages in the selection process differed across regions and types of population points. In the twelve self-representing PSUs, a two-stage procedure was used: selecting census enumeration districts and selecting households.[9] In the urban locations of the non-self-representing PSUs, there were four stages: selecting the two PSUs (where there were more than two units in a stratum); selecting the urban population points (if there were more than enough from which to draw the necessary number); selecting the census enumeration districts, and selecting households. In rural areas of the NSR PSUs, three stages were used: selecting the two PSUs (where there were more than two); selecting the villages; and selecting the households.

As data become available from the survey, this report will be expanded to discuss the representativeness of the actual sample which we obtained in the field, as well as the design effects of our sample design.[10]

---

[9]Since these units were selected with certainty, their selection is not actually a stage from the viewpoint of mathematical statistics. Also, bear in mind that stratification per se is not a stage either. Thus, in these twelve units, census enumeration districts are technically the PSUs; households are the SSUs. Individuals do not count as a stage since all individuals within each household were selected with certainty (though some may have declined to participate).

[10]Direct questions to Dr. Michael Swafford, Paragon Research International, Inc. 511 Fairfax Avenue, Nashville, TN 37212. (615) 383-7733. 385-9761 (fax). Internet: SWAFFOMS@CTRVAX.VANDERBILT.EDU.

Table 1

DISTRIBUTION OF HOUSEHOLDS IN SELF-REPRESENTING RAIONS

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
OKTIABRSKII RAION	100
PERVOMAISKII RAION	84
SVERLOVSKII RAION	<u>108</u>
TOTAL FOR SELF-REPRESENTING RAIONS	718

Abbreviations for Oblasts

DA	DZALAL-ABAD	IK	ISSYK-KUL'SKAYA
OSH	OSH	N	NARYINSKAYA
T	TALASSKAYA	CHU	CHUISKAYA

Table 2

SELECTION OF NON-SELF-REPRESENTING RAIONS

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
Tonskii (IK)	10,233	
Alaiskii (OSH)	11,357	
DZHETI-OGUZSKII (IK)	14,083	95
Toktogul'skii (DA)+Kara-Kul'	17,901	

STRATUM 2 (166)

Manasskii (TAL)	5,054	
Kara-Burinskii (Kirovskii) (TAL)	7,878	
BAKAI-ATINSKII		
Leninpol'skii (TAL)	8,584	83
Talasskii (TAL)	9,104	
Kochkorskii (N)	9,295	
Dzhumgal'skii (N)	9,360	
BATKENSKII (OSH)	9,457	83
Kara-Kul'dzhinskii		
Sovietskii (OSH)	10,496	

STRATUM 3 (153)

Tiupskii (IK)	11,066	
AK-SUISKII (IK)	11,487	76
Aksyiskii (DA) + Tash-Kumyr	18,331	
NAUKATSKII (OSH)	24,440	77

STRATUM 4 (213)

Chuiskii (CHU)	10,459	
PANFILOVSKII (CHU)	12,734	106
Issyk-Kul'skii (IK) + Cholpon Ata	14,702	
Moskovskii (CHU)	20,739	
SOKULUKSKII (CHU)	32,504	107

STRATUM 5 (208)

Kalininskii (CHU)	11,885	
ISSYK-ATINSKII (CHU)	12,356	104
Keminskii (CHU	16,085	
Kantskii (CHU)	20,043	
ADAMEDIVSKII (CHU)	28,371	104

9

STRATUM 6 (140)

Missing Name		
ARAVANSKII (OSH)	12,070	70
Bazar-Korgonskii (DA)	14,926	
NOOKENTSKII (Leninskii)		
(DA)+MALI SAI	23,438	70

STRATUM 7 (148)

SUZAKSKII (DA) + KOK-YANGAK	25,501	74
KARA-SUISKII (OSH) + KARA-SUU	37,782	74

Stratum 8 (168)

Liailiakskii (OSH) + SULIUKTA	15,980	84
UZGENSKII (OSH)+UZGEN	27,365	84
KADAMZHAISKII (FRUNZENSKII)		
(OSH)+KYZYL-KIA	28,100	

**TOTAL IN NON-SELF
REPRESENTING RAIONS**   1,385