Survey Sample Design Report for the project "Coping with Climate Shocks in Mongolia"

anonymized version

prepared by Thomas Otter and the German Institute for Economic Research (DIW Berlin)

April 27, 2012

Project Summary

The project investigates ways in which climate shocks impact households' socio-economic vulnerability and poverty. In doing so, the project accounts for the interconnected economic, social, and environmental causes and effects of climate change on vulnerability and poverty. Moreover, the project explores how risk-management strategies can lead to resilience. The case of Mongolia will serve to study processes of larger relevance concerning the perception and expression of climate risks and their role as a trigger for distress migration. From a theoretical point of view, the project aims to understand the channels through which climate shocks interrelate with rural livelihoods. This objective is implemented by employing different conceptualizations of shocks (endogenous and exogenous), different units of analysis (individual, household, groups of co-residing herders, and settlements), and different time horizons (short, medium, and long). From a methodological point of view, the objective of the project is both to contribute to the development of a quantitative database and to advance applied empirical methods. Together with a Mongolian partner, the National Statistical Office of Mongolia, three yearly waves of household panel data will be collected. The project will outline policy implications with respect to strengthening the long-term resilience of households against climate shocks and facilitating the return of distress migrants to the rural economy.

Objective of Survey

In the project, funded through a research grant from BMBF,[1] researchers at DIW Berlin aim at exploring how households in Western Mongolia cope with extreme climate and socio-economic shocks. One of the key elements of the project consists of collecting and analyzing household panel data from households living in areas that are vulnerable to zuds[2] over three consecutive years. DIW Berlin established a long-term cooperation with the NSO[3] to carry out the collection and processing of household panel data. DIW, however, under the guidance of international consultants is in charge of providing a draft sample design, which serves a discussion inputs for a dialogue with the NSO. The final sample design will be the product of this discussion process. This report presents the final sample design aspects and characteristics, already agreed upon by DIW and NSO through discussions hold in Ulan Baatar in March and April 2012.

Sample Design

The household panel survey covers sample households in the three aimags of Govi-Altai, Zavkhan, and Uvs. The survey is based on a multi-stage sample design in order to ensure that the sample of households is representative for the three strata (aimag centers, sum centers, rural areas).

The population and housing census of 2010 has been used as sampling frame. The NSO made geographic and population data from the 2010 census available for Govi-Altai, Zavkhan, and Uvs for the purpose of drawing a sampling frame.[4]

In a first sampling step, the three aimags are subdivided into mutually exclusive strata of aimag centers (urban areas), soum centers and rural areas (both considered as rural areas). In the second step, in all strata the Enumeration Area (EA) is used as Primary Sampling Unit (PSA). Ten different random draws of the required number of EAs have been generated out of which NSO together with DIW selected the nine different sets (urban, soum center and rural area per three different aimags) which optimize the trade off between geographic coverage and travel costs. In a third sampling step, inside each EA households will be selected again randomly, however with fixed cluster size.

[1] Federal Ministry for Education and Research, Germany

[2] A Mongolian term for an extremely snowy winter in which livestock are unable to find fodder through the snow cover, and large numbers of animals die due to starvation and the cold. The term is also used for other meteorological conditions, especially in winter that make livestock grazing impossible.

[3] National Statistical Office, Mongolia

[4] Variables made available: Aimag code, Soum code, Bag code, Strata, EA code, N households, n population, n herders.

Sample Size Estimation

The following general sample size formula is the basic approach for our estimation which for the specific calculation of the sample size for our study (number of households to be interviewed) is adjusted using the equation (b) below for finite population.

Sample Size - Infinite Population

$$\text{(a)} \qquad SS = \frac{Z^2 \times (p) \times (1 - p)}{C^2}$$

SS = Sample Size
Z = Z-value for two sided test
P = Percentage of population picking a choice
C = Confidence interval, expressed as decimal

Sample Size – Finite Population

$$\text{(b)} \qquad \text{New SS} = \frac{SS}{(1 + (\frac{SS - 1}{Pop}))}$$

Pop = Population

Cluster Sampling

A fixed number of 8 interviews per EA (cluster) was established. Literature usually proposes the implementation of 8 interviews per EA for fixed size clusters. This same number has been adopted already by NSO in the implementation of the 2002 household income and expenditure survey.

Using fixed cluster sizes offers several advantages. During the statistical and econometric analysis sampling errors (which bias results) can be controlled more easily. At the same time a fixed cluster size assures that the total number of individuals in each group is approximately the same (an efficiency consideration); and controls for the possibility that the number of individuals in a randomization unit reflects existing within-cluster dynamics or other factors that are potentially related to outcome. [5]

[5] For example in trials conducted in developing countries, larger villages may have better health outcomes simply because they are located closer to central health facilities. However other factors may also be strong candidates for matching, depending on the main questions of interest. For example, failure to control for

Design effects

Design effects provides a correction for the loss of sampling efficiency, resulting from the use of cluster sampling as opposed to simple random sampling. In practice this correction is included in our estimation via the use of two sided test Z values.

A sample design effect D simply interpreted as the factor by which the sample size for a cluster sample would have to be increased in order to produce survey estimates with the same precision as a simple random sample.

The magnitude of D depends on two factors: (1) the degree of similarity or homogeneity of elements within clusters, and (2) the number of sample elements to be taken from each cluster. The initial factor, the homogeneity of elements within clusters, is a population characteristic over which the survey taker has no control. Prior methodological research indicates that most population characteristics tend to cluster, and thus the prudent course is to assume that some degree of homogeneity within clusters exists. The second parameter, the number of elementary units chosen per cluster, is largely within the control of the survey taker and is an important consideration in the sample design for any survey (see below for further discussion).

Ideally, an estimate of D for the indicators of interest could be obtained from prior surveys in any given setting. Short of this, “typical” values from surveys conducted elsewhere can be used. Since our panel survey will recollect a huge set of different indicators which have different associated specific D values, a “default” value D = 2.0 is used. For analytical purposes most statistical packages correct in their estimation procedures for design effects (e.g. the “svy” command in Stata). A value of D = 2.0 has proved to be efficient in cluster sizes < 20, which is our case (cluster size of 8).

Associated Standard Errors

The implemented sampling strategy allows to achieve statistically significant results ($p<0.05$) with a standard error of 2.29 for the entire survey and 2.8 for urban and for rural areas. Statistically significant results will be achieved at all twelve survey layers (see Table 1 below).[6] The associated standard errors for the different layers are shown in Table 1 below in the three last columns. Naturally, result’s accuracy not only depends on the sample design but as well on the percentage of respondents that pick a particular answer. If 99% of respondents say "Yes" and 1% says "No," the chances of error are remote, irrespective of sample size. However, if the percentages are 51% and 49% the chances of error are much greater. It is

regional differences in socioeconomic status led to interpretational difficulties in a breast cancer screening trial (Alexander et al., 1989).

[6] Entire survey, urban, rural, Zavkhan aimag, Zavkhan urban, Zavkhan rural, Govi-Altai aimag, Govi-Altai urban, Govi-Altai rural, Uva aimag, Uvs urban, Uvs rural.

easier to be sure of extreme answers than of middle-of-the-road ones. Hence Table 1 below shows associated errors for 99, 75 and 50 per cent of response concentration.[7]

Table 1 shows as well the preliminary associated weighting factors. In average, each interviewed household represent roughly 30 households in real life, with some 20 in urban and some 40 households in rural areas. Be aware the final weighting factors will be estimated based on the real data recollected after each wave of the panel survey.

Table 1: Sampling framework

Layers	Aimag	N EAS to be visited	N households to be interviewed	Associated weighting factors	Associated standard errors per response share		
					99%	75%	50%
1	**Zavkhan**	**74**	**588**	**31.5884**	0.79	3.44	3.98
2	Urban	36	288	16.7951	1.11	4.85	5.60
3	Rural	38	300	45.7900	1.11	4.85	5.60
4	**Govi-Altai**	**73**	**583**	**24.9760**	0.79	3.44	3.98
5	Urban	36	286	15.4685	1.12	4.85	5.61
6	Rural	37	297	34.1313	1.11	4.85	5.61
7	**Uvs**	**74**	**592**	**31.4882**	0.79	3.43	3.96
8	Urban	37	294	23.9184	1.11	4.85	5.60
9	Rural	37	298	38.9564	1.12	4.85	5.60
10	**TOTAL**	**221**	**1,763**	**29.3681**	0.46	1.99	2.29
11	Urban	109	868	18.7707	0.64	2.80	3.24
12	Rural	112	895	39.6458	0.64	2.80	3.23

Source: Own estimates based on Census 2010 data provided by NSO.

One final set of EAs has been selected in discussion between NSO and DIW Berlin, out of a set of 10 different random draws. The selection criteria were based on the objective of minimize trade-offs between the required geographic coverage and travel costs.[8] Selected EA's ensure that there are sufficient households in each cluster to be selected for the interviews. In the possible case that the number of available households drops in summer

[7] Holding the level of statistical significance constant at 95%, for smaller subsample areas (layers) naturally the standard error increases. A second source for responses accuracy is the response frequency (concentration of responses on few or single response options. The last three columns of Table 1 show standard errors for 99, 75 and 50% of response concentration. The 50% option can be considered as "worst case scenario". Even in this case standard errors for the entire survey, urban and rural areas and entire aimags are still below 4.0, which can usually still be considered as acceptable. Only splitting up aimag results again into urban and rural increases standard errors for these subareas to levels > 5.0. However, this is only the case for low levels of response concentration (50% and less). In the same way response concentration increases for sub-aimag urban and rural area results, standard errors will go down. For response concentration levels of 80% in aimag sub areas (urban and rural) standard errors drop already below 4.5. An error of 4.0 is reached at 85% of response concentration. Achieving these levels of response concentration can be expected as possible, considering the homogeneity of living conditions specifically in rural areas. A second option of "increasing" response concentration for the results assessment is the grouping (summing up) of different response options of the questionnaire (this is for assessment purposes only, for the implementation of the interviews of response options will have to be considered exactly the way they are expressed in the questionnaire.

[8] All proposed sets of EAs fulfil fully the criteria of random selection required for the simple design.

months due to herding activities, a rule for finding alternative EAs and households has been established (see below in Implementation Instructions).

Sample Attrition

One key aspect of the household panel survey is to trace households who move across strata between panel waves (e.g. households who moved from the countryside to an aimag center or vice versa). Ignoring the movements of households who left the original enumeration area would otherwise result in sample attrition, that is, the sample of households would get smaller with every survey round. Other reasons for attrition are household dissolution, foreign migration, or households' unwillingness to respond – these issues must be minimized as well. Reducing attrition requires particular effort from the interviewers and NSO project managers. First, information regarding the current location and identification of households must be recorded with particular care in order to allow following households over time. Second, the interviewers must be skillful and innovative in acquiring information on the new location where a household has moved, for example by asking neighbors or administrative posts. However, in order to keep costs manageable, only those households will be traced who stay within the three survey aimags of Govi-Altai, Zavkhan, and Uvs. Households moving to Ulan Bator will also be traced and interviewed in Ulan Bator. If a household is known to have moved to a different aimag other than Govi-Altai, Zavkhan, Uvs, and Ulan Bator, it will not be followed.

Timing of Data Collection

A continuous approach of data collection will be applied. This involves conducting interviews throughout the year, with in average 145 households interviewed every month during the entire survey period of three years. In order to ensure the sample is representative across aimags, strata, and seasons, a pre-determined schedule of interviews must be strictly followed. Every sample household will be interviewed precisely every 12 months. For example, if the first interview of a given household takes place in December 2012, the second and third panel interviews with the same household are scheduled for December 2013 and December 2014, respectively. The data will also be entered, processed, and cleaned continuously. Table 3 below shows the timely distribution of EAs to be visited and number of interviews to be implemented per panel wave.

Table 3: Panel survey implementation schedule

EAS to be visited														
			MONTH											Total
EAs to be visited disaggregated in Aimag, Soum and Rural			**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**	**M11**	**M12**	
Zavkhan		***74***	***8***	***5***	***7***	***5***	***7***	***5***	***8***	***5***	***7***	***5***	***7***	***5***	***74***
Urban	Aimag Center	36	3	3	3	3	3	3	3	3	3	3	3	3	36
	Soum Center	8	2	0	1	0	1	0	2	0	1	0	1	0	8
Rural	Rural	30	3	2	3	2	3	2	3	2	3	2	3	2	30
Govi-Altai		***73***	***7***	***5***	***8***	***5***	***7***	***6***	***7***	***5***	***7***	***5***	***6***	***5***	***73***
Urban	Aimag Center	36	3	3	3	3	3	3	3	3	3	3	3	3	36
	Soum Center	10	2	0	2	0	2	0	2	0	1	0	1	0	10
Rural	Rural	27	2	2	3	2	2	3	2	2	3	2	2	2	27
Uvs		***74***	***7***	***5***	***8***	***5***	***7***	***7***	***7***	***5***	***7***	***5***	***6***	***5***	***74***
Urban	Aimag Center	37	3	3	3	3	3	4	3	3	3	3	3	3	37
	Soum Center	10	2	0	2	0	2	0	2	0	1	0	1	0	10
Rural	Rural	27	2	2	3	2	2	3	2	2	3	2	2	2	27
TOTAL		**221**	**22**	**15**	**23**	**15**	**21**	**18**	**22**	**15**	**21**	**15**	**19**	**15**	**221**
Urban	Aimag Center	109	9	9	9	9	9	10	9	9	9	9	9	9	109
	Soum Center	28	6	0	5	0	5	0	6	0	3	0	3	0	28
Rural	Rural	84	7	6	9	6	7	8	7	6	9	6	7	6	84

Number of household interviews														
			MONTH											Total
EAs to be visited dissagregated in Aimag, Soum and Rural			**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**	**M11**	**M12**	
Zavkhan		**592**	***64***	***40***	***56***	***40***	***56***	***40***	***64***	***40***	***56***	***40***	***56***	***40***	***592***
Urban	Aimag Center	288	24	24	24	24	24	24	24	24	24	24	24	24	288
	Soum Center	64	16	0	8	0	8	0	16	0	8	0	8	0	64
Rural	Rural	240	24	16	24	16	24	16	24	16	24	16	24	16	240
Govi-Altai		**582**	***56***	***40***	***64***	***40***	***56***	***48***	***56***	***40***	***56***	***40***	***48***	***40***	***584***
Urban	Aimag Center	286	24	24	24	24	24	24	24	24	24	24	24	24	288
	Soum Center	80	16	0	16	0	16	0	16	0	8	0	8	0	80
Rural	Rural	216	16	16	24	16	16	24	16	16	24	16	16	16	216
Uvs		**590**	***56***	***40***	***64***	***40***	***56***	***56***	***56***	***40***	***56***	***40***	***48***	***40***	***592***
Urban	Aimag Center	294	24	24	24	24	24	32	24	24	24	24	24	24	296
	Soum Center	80	16	0	16	0	16	0	16	0	8	0	8	0	80
Rural	Rural	216	16	16	24	16	16	24	16	16	24	16	16	16	216
TOTAL		**1764**	**176**	**120**	**184**	**120**	**168**	**144**	**176**	**120**	**168**	**120**	**152**	**120**	**1768**
Urban	Aimag Center	868	72	72	72	72	72	80	72	72	72	72	72	72	872
	Soum Center	224	48	0	40	0	40	0	48	0	24	0	24	0	224
Rural	Rural	672	56	48	72	48	56	64	56	48	72	48	56	48	672

Implementation instructions

This chapter reflects in form of questions and answers some points discussed between NSO, DIW and the author of this report.

1) How does the NSO select the actual households in practice?

Randomly, this is that interviewers within each EA look for the first 8 households where a household head or other qualified information source is available for the interview and willing to participate in consecutive waves of interviews (every 12 months)

2) Do we choose households from EAS which are currently there or which were there at the times of the last census in 2010?

Bearing in mind that that census was conducted in winter, there may be EAS now which contain no households, especially not during the summer months as households may have moved. There is some data from early 2012 which shows that there are some population movements compared to 2010. Since the first selection criteria of households is random criteria we definitely choose households which are currently there. In case and EA is selected where are not sufficient households (8) for interviewing, I suggest that one of the direct neighbor EAs which provides sufficient households should be selected.

3) Generally, and on a related point, do we adjust the sample size per EAS in light of the new, updated estimates of the EAS population sizes?

No need for adjustment. Our level of significance is not at EA level. As long as we have sufficient number of EAs per layer and have 8 interviews per EA, we are fine for the field work (data collection). Adjustments for numeric results for the analysis (e.g. changed population size per aimag) will be managed with weighting factors.

4) Replacement of households with incomplete questionnaires

In case a household is willing to participate in the survey but can not fully complete the questionnaire (e.g. because of lack of time or non availability of a well informed interview partner such as the household head) please visit the household a second time. In case the questionnaire can not be completed in two visits please move to the geographically closest neighbour household in the same EA (or in the next geographic neighbour EA in case there are no more households available in the same EA).

5) Randomization of households within each EA

The total number of households in each EA is known on the EA map assign a number to each household beginning on the top left (North-East) of the map. Make a random draw out of the numbers 1 to 27 (in case the EA includes overall 27 households). You might have drawn number 8. In this case figure out which one is the household number 8 on your map. This is your first household to be interviewed. In case there is no qualified interview partner available or the household is not willing to participate move to the next geographic neighbour household to your right.

6) Campsite interviews

For rural areas (only true rural areas, not soum centers), conduct interviews with ALL households living in one campsite (khot ail). In case one enumerator already completed 7 household interviews in a given EA and approaching the 8th household, it is a campsite shared by 4 households. In this and similar cases the enumerator should anyway interview all 4 households (small response distribution biases created by this rule can be corrected adjusting weighting factors). In case the final number of interviews in one EA exceeds the number of 8 the number of interviews in any neighbor EA can be reduced accordingly.

Instructions to Enumerators on how to Select Households

1) How to select households randomly within each EA

In each enumeration area (EA), the enumerator has to conduct interviews with 8 households. The enumerator has to select these 8 households randomly. This means that the enumerator has to conduct interviews with the <u>first 8 households</u> he/she encounters in a given EA in which the household head (or another qualified person to respond to the questions) is available for the interview and willing to participate in 3 consecutive interviews (every 12 months). In case there is no qualified interview partner available or the household is not willing to participate the enumerator has to approach a new household in the same EA.

Random selection means that the enumerator is <u>not</u> allowed to select households for interviews based on the households' characteristics (such as rich households, poor households, or small households). Instead, the enumerator has to conduct interviews with the first 8 households he/she encounters, irrespective of the characteristics of these 8 households. Also, it is <u>not</u> allowed that the enumerator passes by one household in a selected EA without conducting an interview.

2) How to proceed if there are less than 8 households in a given rural EA

Each EA selected for the survey generally comprises of more than 8 households. Hence, it should be no problem for the enumerators to find at least 8 households in a given EA.

Yet, in rural areas, some herders go on seasonal movements during the summer. In case an enumerator does not find 8 households in a given EA, the enumerator has to contact the NSO project manager at the NSO headquarter in UB and ask for advice. The NSO project manager will decide where the enumerator has to conduct the remaining interviews based on the following rule:

- If the problematic EA is inside a bag where there is at least one more EA selected, the enumerator has to go to that EA and conduct all remaining interviews (in addition to the 8 interviews scheduled for that EA). Note that the total number of interviews conducted in this EA is larger than 8. In the questionnaire section "Household Identification" the enumerator has to write "99" for question 5 (Enumeration Area Code).
- If the problematic EA is inside a bag where there is no other EA selected, the enumerator has to go to a new EA that is neighboring to the original EA and conduct all remaining interviews. In the questionnaire section "Household Identification" the enumerator has to write "99" for question 5 (Enumeration Area Code).

3) Conducting interviews with all households sharing one khot ail

In rural areas (only true rural areas, not soum centers), the enumerator has to conduct interviews with <u>all</u> households sharing one campsite (khot ail). For instance, if one given khot ail consists of he household of a wealthy herder, the household of his son, and the household of a hired herder working for the wealthy herder, the enumerators has to conduct interviews (and fill in separate questionnaire forms) for each of the 3 households. Moreover, the enumerator has to fill in the following information in the Section "Household Identification" of the questionnaire:

- Number of households living in this campsite/khot ail
- Number of households interviewed in this campsite/khot ail
- Name of the head of household living in this campsite/khot ail
- The reason why a household in this campsite/khot ail was not interviewed.

Remember that 8 household interviews have to be conducted in each EA. Imagine a situation in which the enumerator already completed 7 household interviews in a given EA. When approaching the 8th household in this EA, it is not a household residing alone, but a campsite shared by 4 households. In this case (and similar cases) the enumerator has to interview all 4 households – even though the total number of

interviews conducted in this EA will be more than 8 (in this case, the number of interviews is 11).

4) Replacing households with incomplete questionnaires

Sometimes a household is willing to participate in the survey, but the questionnaire cannot be completed. For instance, the respondent may be short in time to complete all sections of the questionnaire or the head of household or another well-informed household member may not be present during the interview. If this is the case, the enumerator has to visit the household for a second time. If the questionnaire cannot be completed in the two visits the enumerator has to approach a new household in the same EA. It is important that 8 complete interviews have to be conducted in each EA.

In each household interview, the most important section is on "Household Identification". The information recorded in this section is needed to locate the households in the second and third wave of the survey. It is the responsibility of the enumerator to ensure that all answers recorded in this section are complete and accurate. It is not allowed that any answer in the section "Household Identification" is missing. In case there is any missing information in this section, the enumerator has to approach a new household in the same EA.

All questionnaire forms (even incomplete forms or forms with missing information on the "Household Identification" section) have to be returned to the team supervisors for data entry.

Implementation of Sampling in Soum Centers and Aimag Centers

Enumerators have to select households for interviews in soum centers and aimag centers according to the following rules:

1. The maps of EAS (showing the location of households as dots) for soum centers and aimag centers are the basis for selecting households in soum centers and aimag centers. The maps are derived from the 2010 census.
2. The enumerator has to assign numbers to all households in a given EA that are present in the EA at the time of the interview. That is, the first household present in the EA at the time of the interview is assigned number 1, the second household present in the EA is assigned number 2, and so on. If a household is shown on the EA map but is not present at the time of the interview, this household should be left out from the numbering.
3. The enumerator has to start numbering households at the south east corner of the EA map. South is the first criterion, east the second criterion. That is, the household located at the most south-eastern corner of the map is assigned number 1.

4. After the numbering exercise is complete, the enumerator has to conduct interviews with 8 households in total. The number of the first household is given by the NSO. For example, in EA X, the enumerator has to conduct the first interview with household number 4. To select the second, third, fourth, …, eighth household, the enumerator has to add a step number to the previous (the first) household. The step number is given by the NSO. For example, in EA X, the step number is 8. The second household to be interviewed is 4+8=12, the third household to be interviewed is 12+8=20, the fourth household to be interviewed is 20+8=28 and so on.

5. If one household is selected to be interviewed but refuses the interview or if the interview is incomplete, the enumerator has to select a replacement household. There is also the possibility that the total number of households in a given EA has changed. For example, there is a possibility that household number 28 is selected to be interviewed, but there are only 27 households living in the EA at the time of the interview. In these cases, the enumerator has to choose replacement households until 8 interviews are conducted in every EA in soum centers and aimag centers. The first replacement household is given by the NSO. The same step number should be used. For example, in EA X, the enumerator has to conduct the first replacement interview with household number 5. The step number is 8. The second replacement household to be interviewed is 5+8=13, the third replacement household to be interviewed is 13+8=21 and so on.

Bibliography

Alexander F., Roberts M.M., Lutz W., Hepburn W. (1989) "Randomization by cluster and the problem of social class bias." *Journal of Epidemiology and Community Health*;43:29-36.

Aliaga, A. and R. Ren, Cluster optimal sample size for demographic and health surveys, ORC Macro, Inc., United States, 2007

Kish, L. (1994). Multipopulation Survey Designs: Five Types with Seven Shared Aspects. International Statistical Review 62, 167-186.

Shackman, G. (2001). Sample size and design effect. http://www.albany.edu/~areilly/albany_asa/confweb01/agenda01.htm

UN-DESA, Statistics Division: Designing Household Survey Samples: Practical Guidelines, NY 2005