EGERTON UNIVERSITY
TEGEMEO INSTITUTE OF AGRICULTURAL POLICY AND DEVELOPMENT

&

MICHIGAN STATE UNIVERSITY

TEGEMEO AGRICULTURAL MONITORING AND POLICY ANALYSIS PROJECT (TAMPA 1)

HOUSEHOLD SURVEY 2000 DATA DOCUMENTATION

Contributors:

Margaret Beaver- Michigan State University
Tom Awuor -Michigan State University
Mary Kiiru- Tegemeo Institute
Bridget Ochieng- Tegemeo Institute
Betty Landan- Tegemeo Institute
Joshua Ariga-University of Arizona
Joseph Owuor- Egerton University

Revised 9/1/2001

Publications

Argwings-Kodhek, G. 1998. Monitoring for Improved Agricultural Sector Policy Making. Paper presented at the Kenya Agricultural Marketing and Policy Analysis Conference, Nairobi, Kenya, Tegemeo Institute of Agricultural Policy and Development, May 12, 1998.

Mose, Lawrence, 1998. Factors Influencing Smallholder Fertilizer Use in Kenya. Paper presented at the Kenya Agricultural Marketing and Policy Analysis Conference, Nairobi, Kenya, Tegemeo Institute of Agricultural Policy and Development, May 12, 1998.

Strasberg, Paul, T.S. Jayne, T. Yamano, Gerald Nyambane, Daniel Karanja, James Nyoro, and Margaret Beaver, 1998. Effects of Agricultural Commercialization on Food Crop Input Use and Productivity in Kenya. Paper presented at the Kenya Agricultural Marketing and Policy Analysis Conference, Nairobi, Kenya, Tegemeo Institute of Agricultural Policy and Development, May 12, 1998.

Kiiru, Mary, B. Landon, B. Ochieng, E. Mghenyi, and T Awuor, 2000. TAMPA Household Survey, Year 2000 Data Collection Manual. Tegemeo Institute of Agricultural Policy and Development. Nairobi, Kenya

Sampling Method

The sampling method used was similar across all the sites and is described below:

1.   Within the designated area of study (considering AEZs and other criteria), all the villages/sub-areas were listed with the help of the administration or chief.

        AEZ, population, and whether the district belonged to the "original" KAMPAP districts (districts where Tegemeo had conducted much research before and had some supplementary data and information on) were some of the key factors in this exercise.

        The first step was to identify the spatial distribution of AEZ in the district. The idea was to capture as much of the diverse conditions as possible in our sampling. From this step we were able to classify certain areas within AEZ with the help of the Ministry of Agriculture officers. Each district was in turn divided into divisions, locations and sub-locations and then villages/wards. From the district level we were able to pick representative divisions with the help of the district officers. I believe that we also took into account the populations and AEZ conditions within these areas to help us select these divisions. Because not all divisions could possibly be visited we picked a random sample of these divisions for further follow-up. These were selected with the idea of incorporating the diversities that were inherent in each district that we visited (a representative sample).

        At the division level, a similar exercise was carried out with the help of the Ministry officials. From here the locations were selected randomly. Then from here to sub-locations and then finally the villages/clusters below.

2.   From this list (and considering the sample size required from the area) a number of villages were randomly selected by picking from the list above.

3.   For the selected villages, and with the help of the administration and key informants, we listed all household units within the village by head of household.

4.   In most cases the list above exceeded the sample size requirements for the area. Accordingly we used the 'universal' KAMPAP sampling technique to select households for interview.

        Universal KAMPAP sampling technique description: Most village elders/chiefs have a pretty comprehensive list of householders' names. Suppose we had a total list of 76 households for a village or cluster from the chief (numbered from 1 to 76). Assume too that all we needed was to interview 12 households from this village. The objective was to randomly select every sixth household to get the 12 we needed (approx 76/12=6). The question is, on a numerical list of 1 to 76 where do you start the selection (is it 1,2,3,4,5 or 6)? We wrote the numbers 1 to 6 on different pieces of paper of similar size, folded and mixed them up. Then we asked a villager or the chief to pick one of these papers and reveal the number. Suppose the number picked is 3; then we proceeded to pick the households starting from the third on the list, i.e 3,9,15,21,27 etc.

5.   It happened that in some areas some of the selected households within a village had household heads who were related by marriage, or some other kinship relationship (though the samples had been selected randomly in the first place). In such instances one could find cousins, brothers, uncles, etc who had bought farms in the same area and over the years subdivided their farms to their kids, etc but all these were clearly separate households with different management styles and approaching their household decisions separately. Relationships among households do not necessarily imply joint decision-making.

6.   In conclusion the samples were as random as was possible and the data should be able to express this random nature despite some pockets here and there of 'relationships', if one may.

Procedure used to replace households not available for interview

Replacement was done when more than 20% of the original households in the cluster (village) could not be interviewed for some reasons. This was based on the rule of thumb. The agreed approach was to get out of the original household, go to the right without crossing the road, count three households - the fourth one becomes a replacement of the original household. If unsuccessful in the fourth household the next household was interviewed (see Enumerator's Manual).

If the respondent was not interviewed, the reason for not conducting the interview was recorded. After the replacement was done, a new household identification number was given on the replacement household. This was necessary so that the 2000 data would match the 1997 data.

Summary of households surveyed

Out of the total survey sample of 1609 households (i.e 1997 & 2000), there are 1446 households (excluding Turkana and Garissa Districts data) that were interviewed. Adding in the Turkana and Garissa households, which was a total of 66, there are a total of 1512 households in the year 2000 survey.

The data for page one of the survey instrument is contained in two files: allhhid00.sav and hhidfinl.sav. The first file (allhhid00.sav) contains all the original households as well as the new households (1609 hhids). The second file (hhidfinl.sav) contains only those households that were interviewed for this 2000 survey (1512 hhids). This file should be used to merge the identifying characteristics to the other files as needed.

Data file descriptions for Rural Household Survey - 2000 - Egerton University - Tegemeo Institute / MSU
Updated - 3/04/2001

Directory structure: - first level subdirectory is called **Kenyahh2000**. There are 4 sudirectories off this directory: anal, arch, doc and tmp.
Under the "arch" directory are 5 directories" augdata, lookup, NewVars, origdata, syntax.

\Kenyahh2000\anal - for analysis files and syntax
\Kenyahh2000\arch
    \Kenyahh2000\arch\augdata - final data files to be used for analysis
    \Kenyahh2000\arch\lookup - lookup data files and syntaxes
    \Kenyahh2000\arch\NewVars - SPSS file where new variables have been created to be used for analysis.
    \Kenyahh2000\arch\origdata - original data files - not to be used for analysis
    \Kenyahh2000\arch\syntax - syntax files (e.g., files to create new SPSS data files with new variables to be used for analysis, files used to
        clean data before placing into "augdata" and other syntax files
\Kenyahh2000\doc - final survey and data documentation files
\Kenyahh2000\incprox
    \Kenyahh2000\incprox\lookup - lookup data files and syntaxes for income proxy work
\Kenyahh2000\tmp - directory used to store temporary files that the analyst does not plan to retain.

Location identifying variables: aez - agricultural ecological zones,
                             aezsmall - aez subdivided into more specific zones
                             zone - habitat zones,
                             prov (province), dist (district), div (division), location (location), subloc (sub-locations),
                             vil (village).

Data Tables:

			Directory: \KenyaHH2000\arch\augdata		
Type of Data	**File name**	**Key variables**	**Number of Cases**	**Computed Variables**	**Comments**
household id	hhidfinl.sav	hhid	1512		all households interviewed, including identifying variables, aez, aezsmall, zone, prov, dist, div, location, subloc, vil
Household information	hh00.sav	hhid	1512		
Inventory of crops raised	incrop00.sav	hhid, crpgwn	18793		If the crop was not a tree crop, the value of numprod was set to -889, otherwise, the value was the number of trees.
Cropping patterns - field	field00.sav	hhid, year, season, field	10678		season: 1 = main, 2 = small year: 4=99/2000
Cropping patterns - crop	croplev.sav	hhid, year, season, field, crop	23277		season: 1 = main, 2 = small year: 4=99/2000
Fertilizer used on the field	fert00.sav	hhid, year, season, field, ferttype, fertunit	8558	fertotal - amount used was converted to kgs using a conversion factor	kgconver - obtained from \lookup\fertconv.sav using "Match Files" command
Expenditure on hired labor	lbrcsh00.sav	hhid, year, season, crop, activity	2981		cash paid for labor
	lbrink00.sav	hhid, year, season, crop, activity	173		in kind paid for labor

	Directory: \KenyaHH2000\arch\augdata				
Type of Data	**File name**	**Key variables**	**Number of Cases**	**Computed Variables**	**Comments**
Cash purchase of fertilizer	tfert00.sav	hhid, ferttype, fertsrce, punit	1688		
Inputs received on credit	input00.sav	hhid, Iinput, CredKind, AmtInput, InpUnit, InpValue	995		
Purchases for home consumption	purch00.sav	hhid, purch	12536	conv1, conv2, conv3, conv: conversion unit. kgqty1, kgqty2, kgqty3, kgqtyt: calculated kgs used per time period. kgtotal - total kgs used by hh	purconv.sav - lookup table to convert to kgs
Livestock revenue	lstsld00.sav	hhid, livecode	5063		livestock in stock June 1999, purchased, price, consumed, born, died, stolen, sold, average price, number as of May 2000, current value
Livestock products	lstprd00.sav	hhid, liveprod	2303		average livestock production in milk, eggs, honey, other products, number of months, amount sold/month, price/unit, buyer (milk only)

	Directory: \KenyaHH2000\arch\augdata				
Type of Data	**File name**	**Key variables**	**Number of Cases**	**Computed Variables**	**Comments**
Household demographics	demog00.sav	hhid, perno	12710		sex, age, relation to head, currently in school, years of schooling, months living at home, year left home, why left, if died cause, engage in business/informal labor or salaried employment.
Informal Income	hhinc00.sav	hhid, perno, activity	1775		activity, each month's earnings; low, average, high earnings gross and cost
Salaried wage / permanent employment	salwg00.sav	hhid, perno, activity	1161		activity, monthly wage, if same throughout year, wage earned per each month if not same
Economic activity	ecact00.sav	hhid, econact	4645		economic activity, order of importance
Household assets	asset00.sav	hhid, item	8133	Totval - total value of items	Item, quantity, value, tot (if could not specific value for 1 item)

Lookup tables: \Kenyahh2000\arch\lookup

Type of Data	File name	Key variables	Number of Cases	Comments
Crop quantity conversion to kgs	cropconv.sav	crop, unit	359	Use this file if you want to convert all harvested/sold units to kgs
Fertilizer quantity conversion to kgs	fertconv.sav	ferttype, fertunit	46	This file was used to standardize the quantity of fertilizer when the file was restructured from field level to field, fertilizer level -fert00.sav
fertilizer purchase price	pricefert.sav	ferttype, punit, dist	759	
purchase quantity conversion to kgs	purconv.sav	purch, unit	54	This file was used to calculate the kgs purchased in purch00.sav
price for all crops	pricconv.sav	crop, zone	489	see syntax file - lvstkprice-lookups00.SPS. If fewer than 20 cases existed, the national price was used, otherwise zonal price was used.
prices for livestock products	lstprdprice.sav	liveprod, zone	23	
livestock purchase price	lvstpprice.sav	livecode, zone	47	
livestock selling price	lvstsprice.sav	livecode, zone	93	
sifted maize meal price	siftedmaizemealprice.sav	dist, div	44	
maize purchase price	maizepurchaseprice.sav	dist	22	

Miscellaneous Notes on the Rural Household Survey 2000
Egerton University - Tegemeo Institute / MSU

Updated - 18/01/2001

Household Numbers

Original household numbers from the 1997 survey range from 1 to 1584 for a total of 1540 households. Replacement household numbers for this survey start at 1622. There are 31 replacement households and 38 new households. The new respondents start numbering at 1585 through 1621 and 1653. From the original (1997) sample, 97 households were not interviewed; 1443 households were interviewed. For this survey, a total of 1512 households were interviewed.

Household numbers not used:
    377
    378
    394 - 404
    1134 - 1154

Population sizes for each of the agricultural ecological zones and samples within each of these zones is documented in \KenyaHh97\arch\wp\popaez2.xls.

Notes on Specific Files

Livestock Inventory file - lstsld00.sav

Note that sheep, goat, chicken, ducks, turkeys, pigs, rabbits and camels were **not** initially asked the questions "Number purchased" (purch) through "Average price when sold" (AvgPric). When almost 30% of the questionnaires had been entered, a suggestion from Tegemeo was made, based on field data, that sheep and goats sold and their average prices should be entered because in some areas goats and sheep were significantly traded. It was decided to ask the information of all the livestock. Thus, the section, which was greyed out from purch through avgprice in the survey instrument is coded as -887 (not required / question introduced later). The data that were collected **before** the decision to add the remainder of the livestock for these two questions, "sold" and "AvgPric", are coded as -887 to indicate that the question had not been asked.

It is not possible to check for balances as at May00 against the transactions throughout the year.   There was no provision to collect information for livestock given in or out as dowry or gifts, however the enumerators wrote them down as a footnote in the questionnaire. Also growth of calves to cows was not counted.

Livestock Product file - lstprd00.sav

In this file "Buyer" was asked only for milk. For the rest of the products, where buyer was not asked, the variable was coded as -887 "Not required".

Honey prices varied greatly from area to area so extreme values were hard to identify. Comparisons were made within the same area for extremes.

Hides and skins were sold in both numbers and kilograms. These prices varied so much because the type of skin/hide was not asked. Some were sheep, others were goats or cows.

Household level file - Hh00.sav

Question 5b - total expenditure on salaried worker for the last 12 months. The value of -889 means that the household had not hired a salaried worker and therefore -889 has been defined as user missing.

Question 16 - It is true that farm gate price of maize varies tremendously within the same village. This could be attributed to the fact that

I)   Harvesting period was not defined (it could range over a 2 month period).
II)   The enumerator could have missed the period *immediately after harvest*.
III)   The enumerator may have recorded only 1 year instead of an average for the last 3 years.
IV)   There could have been some bias in answering the question- net buyers could have given high prices while net sellers could have given low prices.

Question 13 and 15 - Variables **bagsstk** (bags of maize in stock from own production before began main season harvest) and **bagstk00** (bags of maize in stock at time of survey June/July 2000) - The survey was conducted in June/July 2000, it is reasonable that a household could have no stock at the time of the survey even though they had some stock before the harvest since the harvest is at different times (e.g. Jul-Aug 1999 for West Kenya, Nov-Dec for Rift Valley).

bagsstk mthfin00 and yrfin00 - Question 13 asked, ''How many bags of maize did you have in stock from your own production just before you began..? There are 8 households that are coded under bagsstk as -889 (did not grow maize). These households are: 434, 629, 631, 675, 676, 683, 684, 687. The 8 respondents who had not planted maize for an unspecified period of time may have considered the above variables invalid to them therefore necessitating N/A. There are 4 other households (670, 689, 904 and 1038) who also did not grow maize. These 4 respondents considered the questions valid to them and hence went ahead and responded to the other questions.

Question 16c - **Acrsoth** - The question asked " what would you do with the land that you choose not to plant maize?- A code 7 - (expand land under maize) was added for those who said they would expand maize acres in Question 16b - How many acres would you allocated to maize?

Question 18a & 18b - **Lowp** (how low would the price of maize need to be before you consider putting some of your maize area under another crop) and **lowlabor** (how low would the price of maize need to be fore you consider reallocating some of your family or hired labor from maize to other income-earning activities such as other crops, dairy, or non-farm jobs): A code was added "-1", to accommodate farmers who would never change their maize allocation irrespective of the price because they do not grow for the market. It was also used for those who would never conceptualize the idea of maize prices coming down to say sh 200/bag since it has never happened in reality. It was even more difficult for some farmers to conceptualize maize being given out free in the market because it has never happened except for relief which was also not reliable.

Question 23d - **Shopchng** (Has location of nearest shopping center changed since 1997) - shopping center was related to the place the respondent bought fertilizer or hybrid seed (i.e. the same place).

1)   If the respondent did not buy fertilizer or seed , then the question on location of shopping center was no longer relevant and a value of **N/A** - not applicable was entered. Then Question 23d1 **fare97** and Question 23e **fare00** were also **N/A**.
2)   If the respondent bought seed, and the location of the shopping center had not changed (**0 = no**), then **fare97** and **fare00** were asked. A value of 0 in these two variables means that the respondent walked to the center.
3)   If the shopping center location had changed (**1 = yes**), then **fare97** and **fare00** were **N/A**, since there was no reference point.
4)   If the respondent bought fertilizer in either 1997 or 2000, then **shopchng, fare97** and **fare00** were **N/A**, since there was no reference point.

Household questions added to the end of the survey after the survey began

Two questions were added after the survey began.

Question 35c was added after the questions on livestock. It asks "What was the average number of lactating cows in the 1999/2000 year?" **laccows**

Question 46 was added to the end of the survey and asked "Over the 1999/2000 season, would you consider your agricultural production to be reflective of a good production year, normal production year or a poor production year? " - **prodyear**

Those households that were not asked these questions were coded as "-887" - question introduced later.

croplev.sav - crop level data.

When a crop was grown in two different forms, e.g. cowpeas and cowpea leaves, the sdtype (seed type) , sqt (seed quantity) and sunit (unit of measure) for one of the crop form could not have a sdtype (seed type), in this example cowpea leaves, and was coded as -889, N/A.

Where the acreage could not be established following the guidelines in the Enumerator Manual, a standard number (plant population) was used to determine the acreage, particularly for perennial crops. For example, 3500 trees/acre was used for determining acreages for tea and 540 trees/acre for coffee. Calculations were done in the field using calculator.

It is important to note that coffee prices reported are for 1997/98 year (obtained from the coffee factories), not for 1999/2000 year. The coffee year was not complete so the farmers could not tell what price they would receive. At the point of the interview, they had only received advanced payments. These prices can be adjusted to 1999/2000 as they become available. Production figures are for 1999/2000.

For crops which were young and not ready for harvesting by the time the survey was done, **-777** (young, not ready for harvesting) was used on hvt (qty harvested).

demog00.sav - demographic file

Hhid and Perno were updated so that the households match correctly across years (12/2003).

Adult equivalent information used in previous surveys.

sex	age	adquiv	sex	age	adquiv
Male	17-39	1.00	Female	17-39	.94
Male	>=40	.85	Female	>=40	.81
Male	6-16	.80	Female	6-16	.75
Male	<6	.30	Female	<6	.25

D6 - Years of schooling. No definition was given for how to calculate the number of years of schooling. Therefore, some enumerators counted the nursery school as part of the number of years of schooling. For children younger than 5 years old the number of years of schooling was set to 0. For children older than 4 the number was not adjusted. The researcher will find several instances where a child may be slightly older than the number of years of schooling, e.g., a child who is 9 years old with 8 years of schooling.

D4 - Relation to head of household. It appears that some of the enumerators coded daughter (son)-in-laws as daughters or sons when in general they should be coded as other relatives. The data were left as coded on the survey instruments.

Hhinc00.sav - informal and business labour activities file

For the variables lgross, agross, hgross - (low, average, high gross), these variables may contain the net cost. Some respondents could not break out the cost so only responded with the net income. Therefore the lcost, acost, hcost variables cannot be used for any reliable analysis on cost.
If a member of the household participates in a particular activity but was away from the home during the interview and could not give details of this informal activity, "-9" - NA, was used. If the specific activity (**activity**) was not know, "-9" was used as the code for the **activity** variable as well as for the rest of the variables for that case.

For cases where informal income was constant, those months were considered average and the income recorded under average gross.

Salwg00.sav. -Salaried wage employment activities file

For those household members who engaged in salaried employment but were not available for the interview, -9 (not available for interview) was used on the variable **activity** if the activity was not known and under **mnwage** where activity was known.

Note: This is for salwg00.sav and hhinc00.sav. Some households could tell for sure that a member of the household was engaged in a particular activity but could not give details, others didn't even know the name of the activity the members engaged in.

Ecact00.sav. - Economic activity file

There were cases where the respondent could not rank the activities the household was engaged in, e.g cases of household head earning money from salary and informal activity but the spouse could not rank the two activities. For these cases -9 (could not rank) was used on the variable **order**.

Another scenario is where a crop was produced but not sold.   In this case, no income was earned from that crop production. The code "-1" - did not earn any income- though produced, was used on the variable **order**.

Purch00.sav. -Purchases file

In the purchases file, tobacco (purch=16) was asked for Turkana only since Tegemeo felt that it formed a major portion of their expenses.

Totqty (total qty in 12 months) was only asked where the quarterly information on purchases could not be obtained, otherwise that variable had -889 - N/A on it. Where nothing was purchased in a particular quarter, 0 was recorded as the value.

Input00.sav - inputs received file

For households who received <u>cash</u> as credit, the total money received was recorded under quantity received (**Amtinput**), 5-numbers- became the Unit of Input (**InpUnit**) and the value of input per unit -**Inpvalue-** was recorded as 1.

C:\docs\kenyahh2000\doc\2000surveyDocumentation_V4.wpd