UGANDA BUREAU OF STATISTICS

UGANDA NATIONAL HOUSEHOLD SURVEY
2002/2003

Data Documentation

1.   Introduction

The Uganda Bureau of Statistics (UBoS) is the central statistical agency in the country. It is mandated to collect, collate, analyse and disseminate statistical information in the country. On a regular basis the Bureau collects and disseminates information on various sectors of the economy like price and inflation statistics, industrial production, population and social statistics; and trade statistics.

The Bureau has so far carried out 8 rounds (see Table 1) of nationally representative surveys since 1988 in its endeavour to collect and update data on a wide range of economic, social and demographic indicators. These household surveys have had varying objectives and scope. Common to all survey round is the socio-economic module. More importantly, the surveys have provided useful information for monitoring welfare in Uganda.

Table 1: Survey round, 1988-2003

Survey Round	Dates	Households covered
Household budget survey (HBS)	Apr. 1989 – Mar. 1990	4,595
Integrated household survey (IHS)	Mar 1992 – Mar. 1993	9,925
Monitoring survey 1 (MS-1)	Aug. 1993 – Feb 1994	4,925
Monitoring survey 2 (MS-2)	Jul. 1994 – Jan 1995	4,925
Monitoring survey 3 (MS-3)	Sep. 1995 – Jun. 1996	5,515
Monitoring survey 4 (MS-4)	Mar. 1997 – Nov. 1997	6,654
Uganda National Household survey 1 (UNHS-1)	Aug. 1999 – Jul. 2000	10,696
Uganda National Household survey 2 (UNHS-2)	May 2002 – Apr. 2003	9,711

UNHS-2[1] is the latest survey conducted from May 2002 to April 2003, excluding the month of September due to the census enumeration in this month. The main objective was to collect high quality and timely data on demographic and socio-economic characteristics of household population for monitoring development performance of the country. Specifically, the survey aimed at:

a) Providing information on the economic characteristics of the population and its economic activity status, that is, the employment, unemployment and underemployment;

b) Generating data for calculating gross output, value added, and other economic indicators required for national accounts purposes;

c) Integrating household socio-economic and community level surveys in the overall survey programme so as to provide an integrated dataset. This will provide an understanding of the mechanisms and effects of various Government programmes and policy measures on comparative basis over time;

d) Meeting special needs of users for the Ministries of Health; Education and Sports; Gender, Labour and Social Development; and other collaborating Institutions, together with Donors and the NGO community so as to monitor the progress of their activities and interventions; and

e) Generating and building social and economic indicators for monitoring the progress made towards social and economic development goals of the Country.

UNHS-2 was a multipurpose survey designed with three modules, namely, socio-economic, labour, informal enterprises and community. The survey covered 9,711 households in 973 communities. The information for the three modules was collected at the same time and from the same enumeration area, making the marching households to their respective communities easier. This survey was also designed to link up with the 2002 National Population and Housing Census for poverty mapping.

This documentation is designed to provide users with the information they need to understand the UNHS-2 dataset and use the data appropriately. The document is organized as follows: Section 2 provides information on

---

the survey instruments used in the collecting data. The sample including designing, size and derivation of population weights is the subject of section 3. Section 4 describes the fieldwork procedures including organization of fieldwork, recruitment and training, and household listings. Survey implementation is presented in section 5. Section 6 discusses data entry and data cleaning prior to dataset structure and use in section 7. Section 8 provides, to some detail, the construction of consumption aggregate used in poverty analysis.

2.   Survey instruments

Several institutions were involved in the planning of the survey including the process of designing the survey instruments, with the Bureau playing the leading role. A stakeholders' workshop was held to take care of all users' data needs. In addition there was a Steering Committee, a form of consultative group involving key data users that were involved in deciding the questionnaire content. The users provided useful comments/suggestions on the modules especially those relevant to their line ministries. The composition of the Steering Committee included UBoS, Institute of Statistics and Applied Economics of Makerere University, World Bank, Economic Policy Research Centre, Population Secretariat, Uganda Manufacturer's Association and Ministries of Finance, Planning and Economic Development; and Gender, Labour and Social Development.

Each of the households was interviewed face-to-face using three modules, namely the socio-economic module, the labour force module and the informal sector module. At a community level, direct interviews were conducted with different relevant authority. All questionnaires include the geo-referencing data on longitude and latitude for every household, primary school and healthcare units.

2.1   Household listings
The purpose of the listing questionnaire was to obtain a comprehensive updated list of all households in the selected Enumeration Area (EA). This was used as a basis for selection of a random sample of households to be interviewed. The cartography map used in the 2002 Population and Housing Census was used to identify the boundaries of the selected enumeration areas.

The listing questionnaire had two broad categories of stratification of households: those with at least one unemployed person and those with a household enterprise that was non-crop farming. The households were categorized into four and from each category simple random sampling was used to select the sample in the proportion shown in Table 2. In case there was no response, a household would be substituted from within the group, and if there was no household in a certain group, it would be substituted with that in the next group. This resulted in about 5% replaced households.

Table 2: Category of households and sample ratios

Category of household	Sample
With any unemployed person and with none crop farming household	1
With any unemployed person but without any non crop farming household enterprise	1
Without any unemployed person but with a non crop farming household enterprise	6
Without any unemployed person and without a non-crop farming household enterprise	2

2.2 Socio-economic questionnaire

The socio-economic questionnaire contained 9 sections; each covered a separate aspect of the household activity. It provides information at individual and household levels. The various sections are described below.

1   Household identification, including geo-referencing codes

2   Household roster including basic information such as sex, age, marital status of everyone living in the household, and survival of parents for children below 18 years.

3   This section has two parts. Part A information related to health seeking behavior of household members such as type of sickness/injury, health care provider, reasons for not seeking care, etc and Part B information relating to AIDS such as source of information, awareness of prevention methods etc.

4   Education and literacy for all household members, including highest level attained, reasons for dropping out of school, ownership and distances for those still in school etc.

5   This section has three parts. Part A - Housing conditions such as number of rooms, main material used for the wall, roof and floor, utilities and source and distance to safe drinking water, Part B provides information on sanitation; and Part C, migration of the household head.

6   Section provides information on the actual household consumption expenditure and made up of four different parts. Part A – food, beverages and tobacco expenses in the last 7 days; Part B, non-durable and frequently purchased services during last 30 days; Part C, semi-durable and durable goods and services during last 365 days; Part D, non-consumption expenditures in the last 365 days.

7   Provides information on the current value of household and enterprise assets

8   Qualitative welfare indicators

9   Non-crop farming enterprise particulars during last 12 months.

2.3 Labour questionnaire

The labour force questionnaire contained 5 sections; each covered a separate aspect of labour activity. The information contained is at individual level for the usual members as well as the regular members present at the time of interview and aged 5 years and above.

1   Provides information on household particulars where individual leave

2   Usual activity status such as economic status during the last 12 months, employment status, industry code, occupation, reasons for not being economically active and any involvement in vocational training.

3   Made up of two parts, A and B. Part A, provides information on the current activity status for all members above 5 years such as hours worked, employment status, industry and occupation status; Part B, provides information on the current economic status for economically active and employed persons such as wages, years of experience, and under employment.

4   Provides detailed information on the unemployed household members such as reasons for not working, previous type of work, sources of assistance etc.

5	Seeks to provide information on working children aged 5-17 years from an economic perspective.

## 2.4	*Informal sector questionnaire*

This module collected information on the non-crop farming household-based enterprises in both rural and urban areas and rural-based small-scale establishments. These are businesses undertaken by households with or without a fixed location. In addition, inputs and outputs of these enterprises for the major items were also collected. The informal enterprises covered included livestock, poultry, bee-keeping, and fishing; forestry; mining, Quarrying and Manufacturing; hotels, lodges and eating places; and trade and services. For each enterprise/establishment the questionnaires contained 6 sections as indicated below:

1	Provides information on the household particulars where the enterprise is located.
2	Basic background information of the enterprise/establishment such year of establishment, sector, ownership, sources of funding, constraints, and performance.
3	Persons engaged and payments made to them during the last 30 days such as employment status, sex, age, wage & salaries both in cash and in-kind.
4	Provides information on the value of other inputs other than labour during the last 30 days.
5	Made of three parts, A, B & C. Provide information on sales of the enterprise products and any income from other than the sale of products.
6	Establishment/enterprise assets for only non-household establishments.

## 2.4	*Community questionnaire*

In addition to the questionnaires at the household level, a community questionnaire was administered in every EA of the sample. It was administered during the same period as the socio-economic and informal sector surveys. Information on a range of community level variables and conditions was collected through interviewing key informants. The questionnaire contained 8 sections as presented below:

1	General community characteristics and identification particulars
2	Provides information on the nearest school and health facilities by ownership such as availability with the community, distance from the centre, means of transport and time it take to get there.
3	Community history and major events including access to and availability of social services namely schools, clinics, outlets for agricultural and non-agricultural produce in 1992, 1996 and 2002.
4	Predominant Land tenure pertaining in the community,
5	Whether the community received the statutory 25 percent Graduated Tax refund,
6	Development projects undertaken at the community level in the recent past, and whether such projects met the needs of the community
7	Provides information on the primary education infrastructure for the most popular and nearest schools. Such information include, name of school, year of establishment, ownership, type of school, number of students/teachers, condition of the buildings, maintenance plus geo-referencing information for each school.
8	Provides detailed information on the most commonly used private and public health facilities, such as availability of qualified doctors, nurses, midwives, availability of antibiotics, malaria drugs, etc. plus geo-referencing information for each facility.

## 3.	Sample

The number of administrative districts has being changing over time, with new districts being created from the existing one. At the time of the survey Uganda was divided into 56 administrative districts. The newly created districts by the time of the survey are shown in Table 3. The survey was designed to gather estimates at the national and regional levels. Estimates that are representative at the level of a district can only be generated for Masaka, Mukono, Wakiso, Mbale, Lira and Mbarara districts. UNHS-2 sampled households in all the districts in Uganda except Pader district and some parts of Kitgum and Gulu due to insurgencies as the time of the survey.

Table 3: New districts

New districts	Previously from district
Wakiso	Mpigi
Kayunga	Mukono
Mayuge	Iganga
Sironko	Mbale
Kaberamaido	Soroti
Yumbe	Arua
Pader	Kitgum
Nakapiripiriti	Moroto
Kamwenge	Kabarole
Kyenjojo	Kabarole
Kanungu	Rukungiri

3.1 Sampling design and sample Size

The sampling design was chosen to fit the purpose of the survey. Stratified two stage sampling was adopted, but with a few refinements such as over-sampling of urban areas, and possibly of some rural areas with concentrated informal sector activity. The sampling frame for selection of first stage units (FSUs) was the list of EAs with the number of households based on cartographic work for the 2002 Population and Housing Census. For selection of the second stage units, which were the households, listing exercise through listing schedules was done in selected EAs.

Each district was a stratum and was divided into rural and urban sub-strata. The urban area was further sub-divided into district town and other urban areas. This deep stratification enabled a better spread and representation of the sample, thereby increasing the efficiency of the estimates. Additionally, the continuity over rounds was maintained to enable pooling of results over rounds, if ever considered necessary. The total number of about 1,000 FSUs was firstly allocated between urban and rural in the proportion of 40:60. Thereafter, the urban and rural sample was generally allocated between the strata in proportion to the number of households with certain adjustments. The allocated sample was selected with probability proportional to number of households. A suitable plan for sub-stratification and selection of households at the listing stage, was introduced to ensure adequate representation of households with at least one unemployed person and an informal sector enterprise activity.

The households were at first divided into 2 groups namely; households with at least one unemployed person and households with no unemployed person. The total 10 sample households in an EA were allocated between the unemployed and employed groups. Half of the sample from the unemployed was selected from households having one or more household enterprises, while the other was selected from the households having no enterprise activity. In case any one of the sub-groups did not exist, the total sample would be allocated to the existing sub-group. For odd sample sizes, the group with household enterprises got preference.

The households to be selected from the group that did not have any unemployed person were sub-stratified by kind of informal sector activity. The allocation between the sub-groups was in proportion to the number of households, with a minimum of 1 from each group.

As explained earlier, the allocation of the total sample between strata, was not strictly proportional to the number of households. Firstly, the urban areas were over-sampled and secondly some areas both in rural and urban were over or under sampled on the basis of degree of concentration of informal sector activity. Another refinement in the design made was to have a balanced independent inter-penetrating network of sub-samples (IPNS), on a quarterly basis to enable studying seasonality of some survey variables, to provide independent quarterly estimates and to eliminate seasonal effects while taking the average over four quarters. As mentioned earlier, the

need for spreading the survey over a 12-month period and balancing the design arose because of inclusion of labour-force and informal sector survey modules in this round. Ugandan experience indicates the presence of seasonality especially in self-employed activities. But this aspect has never been studied precisely in the past and users needed these data from this round. It is important to note that all the ten households randomly selected from each EA were not interviewed in the same calendar month.

3.2 Estimation procedure and calculation of weights/multipliers

Estimates were built initially at the basic stratum-level and then added over strata to obtain the needed final estimates. As an illustration, in order to derive the necessary formulae for estimation, an estimate of the total, Y_R, say total number of gainfully employed persons in the Rth stratum (say in the rural areas of a given district) can be obtained using the following steps:

First Step: There are nine sub-strata of households by employed/unemployed and kind of informal enterprises formed in each selected EA. The first step will be to get an estimate of gainfully employed in the ith selected EA by adding all the estimates of all the sub-strata by using the formula given in equation (1).

(1) $$\hat{Y}_{Ri} = \sum_{l=1}^{9} \frac{N_l}{n_l} \sum_{k=1}^{n_l} y_{ilk}$$

where:

y_{ilk} = gainfully employed persons in the kth household in the lth sub-stratum of the ith EA;

N_l = total number of households in the lth sub-stratum; and

n_l = number of sample households in the lth sub-stratum.

Second step: The next step will be to build estimates for the district rural ($\hat{Y}_R$) by deriving estimates from each sample EA and averaging over all sample EAs using the formula given in equation (2).

(2) $$\hat{Y}_R = \frac{1}{m} \sum_{i=1}^{m} \frac{h_o}{h_i} \sum_{l=1}^{9} \frac{N_l}{n_l} \sum_{k=1}^{n_l} y_{ilk}$$

where:

m = number of sample EAs selected in the district rural stratum;

h_o = total number of households in the district – rural stratum as per the latest available records used for sample selection; and

h_i = total number of households in the ith sample EA as per the latest available records.

Estimates of district total can be got by adding estimates over two or three strata as the case may be. Similarly, regional and national estimates can be prepared by simple additions. For details the stratum-wise distribution of allocated sample of first stage units (FSUs) for UNHS 2002/03 see UBoS, Administrative Report. The FSUs were the EAs.

4. Fieldwork

4.1 Organisation of fieldwork

The Survey staff comprised of a total of 15 field teams. Fieldwork was undertaken with the use of centrally recruited field teams whereby work in the sampled areas was programmed from headquarters. There are four statistical regions in Uganda namely Central, Eastern, Northern, and Western. All the teams were recruited based

on the languages most prevalent in each region. Four teams were recruited for each region except for Northern region which had three.

Each field team was composed of one supervisor, four enumerators, and one driver. There was no gender bias and these supervisors and interviewers could be either male or female. For instance, out of 75 enumerators and supervisors, 30% were female. Each field supervisor was responsible for one team of Interviewers.

During each round of fieldwork, each team supervisor had to write a report detailing what transpired on the respective trip. It included technical issues concerning the questionnaires where he/she could not take outright decision, coverage in terms of the EAs allocated, and the administrative issues.

The enumerator had to make a comprehensive list of all households in the enumeration area, select the households to be interviewed by sampling, administer the socio-economic, labour, and informal sector survey questionnaires. On the other hand, the team supervisor was tasked to make appointments with Local Council 1 (LC1) officials for the enumeration of their village, equitably distribute work amongst team members, check through the sampling process done by the enumerator, and check through questionnaires administered by the enumerators. He/she was also supposed to conduct an interview using the community questionnaire with LC1 officials and opinion leaders as respondents, interview the institution heads i.e. schools and health units most used by the community. Where applicable, he/she would also at times assist the enumerators in administering the household questionnaires. Repeated visits were made to those households where no 'well informed' member to be interviewed was found. In addition to the field teams, the survey management team as well as Bureau Staff at the level of officers and above undertook field supervision. On average, each team was assigned two days to complete an enumeration area. An allowance of one day was allocated to teams that had to travel very long distances from the UBoS head office. For details on the number of field teams dispatched and the number of field days allocated for each survey month refer to the UBoS, Administrative Report, 2003.

In the central office at the UBoS headquarters there were editing clerks who performed the data scrutiny before the data were sent for data entry. The field staff who did not go for field work during the respective field trips acted as the edit clerks.

4.2   Recruitment and training

The Bureau recruited survey field staff mainly from a pool of experienced staff who had participated in previous surveys done by the same Institution and were thus familiar with large-scale survey methodology and procedures, and were experienced interviewers or team leaders.

Training was done at a central venue in Entebbe by UBOS technical staff, assisted by the Survey Design Consultant, Mr. S. K. Gupta. The technical staff were UBOS statisticians in the Household Survey Section, and were assisted by UBOS support staff. Training for both supervisors was done at the same time for a period of 10 days (interrupted by a weekend). This aimed at transferring the exact message to both groups. This involved classroom instructions, conducting mock interviews among trainees, and a field practice session where the trainees went to households not in the sample to polish their interviewing skills.

Among other things, training involved imparting skills on basic map reading, on how to exhaustively list all households in a given village, how to select a random sample selected village, and on general interviewing skills. In particular, the trainers paid special attention to see that the interviewers understood all the concepts and definitions relevant to the survey. Classroom exercises were conducted to make certain that the respondents had grasped the skills.

5.   Survey implementation

Schedule of activities: The whole exercise of UNHS-2 was undertaken through a work schedule that covered a period from August 2001 to December 2003 (see Table 4).

Table 4: Schedule of survey activities, 2001-2003

Activity	Dates
Stakeholders workshop	
Pre-testing survey	Dec. 2001
Training of enumerators	April 2002
Training of supervisors	April 2002
Actual fieldwork	May 2002 – Apr. 2003
Data entry	Jun. 2002 – May 2003
Data cleaning	Jun. 2003 – Jul. 2003
Poverty analysis	Aug. 2003 – Oct. 2003
Report writing	Nov. 2003
Dissemination workshop	November 2003

Initially, the survey had targeted a sample of about 1,000 enumeration areas and 10 households from each of these enumeration areas. However, due to insecurity in Pader and some parts of Gulu and Kigtum districts were not covered, reducing the sample to 973 enumeration areas and 9,711 households as shown in Table 5.

Table 5: Distribution of EAs and Households Surveyed in each District, UNHS 2002/03

District Code	Name	No. of EAs	No. of Households	District Code	Name	No. of EAs	No. of Households
	Central				**Northern**		
101	Kalangala	8	80	301	Adjuman	8	80
102	Kampala	28	275	302	Apac	20	200
103	Kiboga	12	120	303	Arua	22	220
104	Luwero	24	240	304	Gulu	17	170
105	Masaka	40	397	305	Kitgum	8	80
106	Mpigi	20	200	306	Kotido	15	150
107	Mubende	24	240	307	Lira	30	300
108	Mukono	32	320	308	Moroto	11	110
109	Nakasongola	8	80	309	Moyo	12	120
110	Rakai	20	200	310	Nebbi	16	160
111	Ssembabule	12	120	311	Nakapiripirit	8	80
112	Kayunga	16	159	312	Yumbe	6	60
113	Wakiso	40	400				
	Eastern				**Western**		
201	Bugiri	16	160	401	Bundibugyo	8	80
202	Busia	16	159	402	Bushenyi	20	200
203	Iganga	24	239	403	Hoima	12	120
204	Jinja	28	279	404	Kabale	24	240
205	Kamuli	20	200	405	Kabarole	24	240
206	Kapchorwa	12	119	406	Kasese	24	240
207	Katakwi	12	120	407	Kibaale	8	80
208	Kumi	12	120	408	Kisoro	12	119
209	Mbale	32	320	409	Masindi	16	160
210	Pallisa	16	160	410	Mbarara	40	397
211	Soroti	20	199	411	Ntungamo	12	120

212	Tororo	24	240	412	Rukungiri	12	120
213	Kaberamaido	8	80	413	Kamwenge	12	119
214	Mayuge	12	120	414	Kanungu	12	120
215	Sironko	16	160	415	Kyenjojo	12	120

6. Data entry and data cleaning

6.1 Data entry

Data entry was done at the UBoS heaquarters from June 2002 to May 2003. A total of 16 data entry operators participated in data entry supervised by two data processing supervisors. There were two shifts, the first running from 8:00 a.m. to 2:00 p.m. and the second from 2:00 p.m. to 8:00 p.m.

CsPro was used to develop the data entry applications; Consistency and range checks were implemented in the CsPro data processing applications. Visual Basic was used to develop data Management Applications. These could automatically convert CsPro flat data files into Ms-Access, which was the final data storage tool. The system developed for data management could coordinate the sampled EAs, interviewed households, and entered households. It could also coordinate entered, compared, cleaned and converted data files. This left the data with very minimal completeness, consistence and data value errors.

Other measures were put in place to minimize errors at the data entry stage. Data entry was done twice, that is, main entry and verification; there was 100 percent verification. A questionnaire could be considered clean after both datasets were perfectly comparable. When an entrant encountered a field problem, a field officer would be called upon to rectify the error.

6.2 Data cleaning

Besides the editing done before data entry, the validation checks inbuilt in the program and double data entry, additional in-depth data cleaning on sections relevant for basic poverty analysis was done by the Economic Policy Research Centre (EPRC). For instance, individual level files were linked together to ensure that the same individual code reported in different sections of the questionnaire and in other modules corresponded to the same individual. Data cleaning on the other sections was done at UBoS headquarters. Any inconsistencies, data entry errors etc found were corrected after checking the original questionnaires. This being a large national household survey there might still be some "inconsistencies", which when found should be reported to the Bureau for clarification.

7. Dataset structure and use

7.1 Data structure and management

The dataset is best understood if used along with the manual of instructions, code list and the questionnaires, which are included on the CD. The questionnaires contain the exact wording of the questions and skip patterns for some questions. They also contain useful information in interpreting the codes. All codes with the exception of employment codes, industry codes, district codes and codes for unit of quantities are contained in the questionnaires. In addition, the population weights must be used to generate results that are valid at national, regional and rural/urban levels. It should be noted that the population weights for the informal sector are different from those used for the other modules.

The dataset, regardless of storage format, is divided into a total of 69 files representing different sections of the questionnaire involved. The socio-economic contains information at individual, household and item levels.

The data from the data entry program were converted into other formats including SAS, STATA, ACCESS and Excel. The description of the data files is given in Table 6. These particulars are also included on the CD.

Table 6: Description of the data files

Filename	Topic	Level of Observation	Max. Observations
Socio-Economic:			
Socio_sec1.*	Identification Particulars	Household	9711
Socio_sec2.*	Personal Characteristics	Individual	52089
Socio_sec3a.*	Health of household members	Individual	50510
Socio_sec3b.*	Information relating to AIDS/HIV	Individual	31430
Socio_sec4.*	Household members' education	Individual	50495
Socio_sec5a.*	Housing Conditions	Household	9711
Socio_sec5b.*	Household conditions	Household	9711
Socio_sec5c1.*	Migration of household head	Household	9711
Socio_sec5c2.*	Places where household has ever lived	Individual	8212
Socio_sec6.*	Persons present in the last 7 days	Household	9711
Socio_sec6a.*	Food, beverages & tobacco	Item	24136
Socio_sec6b.*	Non durable goods & frequently purchased services	Item	99519
Socio_sec6c.*	Semi durable & durable goods & services	Item	81694
Socio_sec6d.*	Non-consumption expenditure	Item	19366
Socio_sec7.*	Household & enterprise assets	Item	71419
Socio_sec8.*	Welfare Indicators	Household	9711
Socio_sec91.*	Household has any non-crop farming enterprise	Enterprise	9711
Socio_sec92.*	Relevant information of the enterprises	Enterprise	6989
Socio_sec93.*	Any enterprise activities in the past 5 years that are no longer running	Enterprise	9711
Socio_sec94.*	Relevant information of enterprises no longer running	Enterprise	1594
Socio_sec95.*	Finishing time	Household	9711
Labour force:			
Labour_sec1a.*	Section 1A: Identification Particulars	Household	39747
Labour_sec2.*	Section 2: Usual Activity Status (for usual members 5 years & above)	Individual	39546
Labour_sec3a.*	Section 3A: Current Activity Status for all Persons aged 5 years & above during the last 7 days	Individual	18081

Filename	Topic	Level of Observation	Max. Observations
Labour_sec3b.*	Section 3B: Current Activity Status for Economically Active-Employed Persons	Individual	21789
Labour_sec4.*	Section 4: Probing Questions for unemployed persons from Section 3A aged 5 years & above	Individual	18890
Labour_sec5.*	Section 5: Economic Activities for Children aged 5-17 years	Individual	9707
Community:			
Comm_sec1.*	Section 1B:Community identification particulars, Section 1C-staff details and survey time		973
Comm_sec2.*	Section 2: Community characteristics		8548
Comm_sec3.*	Section 3: Community history and major events		15519
Comm_sec4.*	Section 4: Land		973
Comm_sec5.*	Section 5: Graduated Tax		973
Comm_sec6.*	Section 6: Community Projects (last 3 years)		16374
Comm_sec7.*	Section 7: Educational – Infrastructure		973
Comm_sec8.*	Section 8: Health infrastructure		
Informal -Trade:			
Trade_sec1.*	Section 1: Identification particulars.		2864
Trade_sec2.*	Section 2: Background information of the enterprise/ establishment.		3820
Trade_sec3.*	Section 3: Persons engaged and payments made to them during the last 30 days.		6895
Trade_sec4.*	Section 4: Value of other inputs other than labour during the last 30 days.		11141
Trade_sec5a.*	Section 5a:Total sales and other incomes during the last 30 days.		1153
Trade_sec5b.*	Section 5b: Incomes received other than sale of products.		2941
Trade_sec6.*	Section 6: Establishment assets.		2864
Trade_sec6b.*			2826
Informal - Hotel:			
Hotels_sec1b.*	section 1b: staff details and survey time		810
Hotels_sec2.*	section 2: background information of the enterprise/ establishment		1163
Hotels_sec3.*	section 3: persons engaged and payments made to them during the last 30 days/month		2773
Hotels_sec4.*	section 4: value of other than labour during the last 30 days		1657
Hotels_sec5.*	section 5: value of receipts, services and goods sold during the last 30 days		1326

Filename	Topic	Level of Observation	Max. Observations
Hotels_sec6.*	section 6: establishment assets (to be filled in for only non-household establishments)		810
Hotels_sec7.*			808
Informal - Forestry:			
Forestry_sec1.*	Section 1: Identification particulars.		230
Forestry_sec2.*	Section 2: Background information of the enterprise/ establishment.		368
Forestry_sec3.*	Section 3: Persons engaged and payments made to the them during the last 30 days.		132
Forestry_sec4.*	Section 4: Value of other inputs other than labour during the last 30 days.		264
Forestry_sec5a.*	Section 5a:Total sales and other incomes during the last 30 days.		8
Forestry_sec5b.*	Section 5b: Incomes received other than sale of products.		162
Forestry_sec6.*	Section 6: Establishment assets.		230
Forestry_sec7.*			227
Informal - Livestock:			
Livestock_sec1.*	Section 1B: staff details and survey time		812
Livestock_sec2.*	Section 2: Background Information of the Enterprise/Establishment		1825
Livestock_sec3.*	Section 3: persons engaged and payments made to them during the last 30 days/month		1646
Livestock_sec4.*	Section 4: value of other inputs other than labour during the last 30 days		1322
Livestock_sec5a.*	Section 5A: total sales poultry and livestock during the last 30 days		726
Livestock_sec5b.*	Section 5B: total sales of livestock, poultry, fish, and bee-keeping products in the last 30 days		59
Livestock_sec5c.*	section 5C: Income received other than sale of products		879
Livestock_sec6.*	Section 6: Establishment assets (to be filled in for only non-household establishments)		812
Livestock_sec6a.*			
Informal - Mining:			
Mining_sec1.*	sec1 : staff details and survey time		1720
Mining_sec2.*	sec2: back ground information of the enterprise/establishment		2739
Mining_sec3.*	sec3: persons engaged and payments made to them during the last 30 days/month		5825
Mining_sec4.*	sec4: value of other inputs other than labour during the last 30 days		2202
Mining_sec5a.*	sec5a: output, transfers, and sales of production during the last 30 days		271

Filename	Topic	Level of Observation	Max. Observations
Mining_sec5b.*	sec5b: income received other than sale of output		752
Mining_sec6.*	sec6: establishment assets (to be filled in for only non household establishments)		1720
Mining_sec6a.*			

Naming conventions:

The questionnaire is the basic guide to the data set. Each section in the questionnaires is numbered and the dataset reflect this structure. For instance, sec1.* contains information from section 1 of the socio-economic questionnaire. When a section is made of more than one part, the parts are also found in the filename. For instance, section 3 is split into two parts, A and B. Within each of the data file, the naming of variables is also based on the questionnaire. It is formed by a combination of letter *s* plus the section number, followed by letter *q* and the number that follow the questions in the questionnaire. Thus, question 4 of section 2 will have the name s2q4. In the case that a question has more than one part, the question name reflects this.

7.2   Merging files

Merging files requires that each unit of observation have a unique identifying code.

Household identifier (**hh**)	aaabcddddee	11-digits
Community identifier (**comm**)	aaabcdddd	9-digits
Individual identifier (**hhpid**)	aaabcddddee f gg	14-digits
Informal enterprise identifier (**hht**)	aaabcddddeeh	12-digits

where:

a: district/stratum code (101-415) and first digit represents region code (1-4)
b: area (1-3)
c: a fill up of zeros (always 0)
d: enumeration area/fsu's, ea (1-2876)
e: household number (1-16)
f: type of person, tid (1-4)[2]
g: personal identification within tid, pid (1-31)
h: type enterprise, enttype (1-3)

Use **hh**, a household identifier, to merge household and individual level information. The same variable can be used to link information household level information with information in the informal sector questionnaire (for only those households reporting a household based enterprise) after dropping the last digit on variable **hht**. The household level files can be linked to the community data files through the constructed variable **comm**, by dropping the last two digits on variable, **hh** in the socio-economic model. This variable is included in socio_sec1 data file. To link individual level files across different sections within/between modules use the constructed unique individual identity variable, **hhpid**. Alternatively sort individual level sections in this order at once, *hh tid pid* OR *hh sid.*

7.3   Missing values

Every effort was made to ensure that the respondents provided the information as requested during the interview. However, as participation in the survey was voluntary there are instances where the respondent(s) deliberately refused to answer or could not provide information on behalf of members who were not present at the time of the interview. For instance a missing code, **222**, has been added to the original dataset.

7.4   Other codes not included in the questionnaires

Variable	Codes	Module
area	1. District town 2. Other urban 3. rural	Socio-economic

---

[2] . For section 2 of the socio-economic theses codes range from 1-7, where codes 5-7 refers to those individuals who have permanently left the households and no data were collected on these individuals elsewhere in the questionnaires.

Type of dwelling	6. hut	

7.5   Other data files

Some of the data files were generated based on the raw household survey data and others are from administrative data. Table 7 presents the details of the files with their respective variables. **Pov0203.*,** contains 16 variables among which is the consumption aggregate used in poverty analysis constructed based on the methodology as presented in section 8; **imputerent.*,** contains 3 variables; and **cpi.*,** contains information on the consumer price indices by category of goods in 1997/98 prices.

Table 7: Variables in constructed data files

Filename	Variable	Description
Pov0203.dta	hh	Household identifier
	region	Region
	stratum	District taken as the stratum
	ea	Enumeration area taken as PSU
	mult	Population multipliers
	urban	Urban dummy
	rmult	Rounded population weights
	regurb	Region by rural/urban dichotomy
	hsize	Total household size, head count
	equiv	Total household size, in adult equivalents
	exdis	District dummy for those districts excluded in UNHS-1
	cpexp30	Monthly household expenditures in market prices after adjusting for regional and intertemporal price variation in 1997/98
	welfare	Derived by dividing cpexp30 by equiv
	hmult	Household weights derived by multiplying rmult and hsize
	spline	Absolute total poverty line in 1997/98 prices
	Poor	Poor dummy 1=poor & 0=non-poor
imputedrent.dta	Hh	Household identifier
	s6bq2	Item code
	imputed	Imputed rent for those with missing rent values
findex03.dta	Sx	
	Findex	Regional food index, 2002/03
	Rname	Region by rural/urban dichotomy, names
	Regurb	Region by rural/urban dichotomy numeric codes
cpi.dta	Year	Year
	Month	Month
	Obs	Observation number
	allitems	Overall CPI in 1997/98 prices
	Food	CPI for food
	beverages	CPI for beverages
	clothing	CPI for clothing
	Rent	CPI for rent etc
	pergoods	CPI for personal goods
	transcom	CPI for transport and communication
	Heeduc	CPI for health and education

7.5   Organization of the CD

The information on the CD is organized in three main folders, namely, data, reports and documentation, details are given below:

Table 8: Organization of information on the CD

Main folder	First level	Second level
Data	There are four folders that are arranged by storage format. These include **STATA, ACCESS, SAS, EXCEL.** Within these folders, data files are arranged by questionnaire as shown in the last column.	Socio
		Labour
		Community
		Hotel
		Livestock
		Mining
		Trade
		Forestry
	The fifth folder is **Constructed data** that contains the consumption aggregate and related files.	There are four folders that are arranged by storage format. These include **STATA, SPSS, EXCEL.**
Documentation	Contains three folders, namely:	
	Manual including item code list	
	Data documentation	
	Questionnaires	Containing five folders by module, that is, socio, community, labour and enterprises plus listing questionnaire
Reports	Contains socio-economic survey report in pdf format.	

8.   Construction of consumption aggregate

Poverty analysis based on the survey data was completed in October 2003. The consumption aggregate is included on the CD to allow researchers/analysts to produce results that are consistent with the officially published poverty figures. Below is a brief documentation of the methodology used and data transformation carried out. In measuring poverty, there are three critical issues: how to measure welfare, how to set the poverty line and how to aggregate over individuals. We address each in turn in sections 8.1-8.3 prior to data transformation in section 8.4.

8.1   Measuring welfare

As stated in the World Development Report 2000/2001, "*poverty is pronounced deprivation in well-being*". Consequently, measuring well-being – or welfare – is the first step in measuring poverty. It is widely recognised that there are many dimensions of well-being. This was one apparent from "Voices of the Poor", a widespread consultation with poor people in developing countries conducted by the World Bank for their World Development Report 2000/2001. From people's responses, a large number of dimensions of well-being were listed ranging from material well-being (including lack of food, shelter, clothing, poor housing) through physical well-being and security to less tangible aspects such as freedom of choice and social well-being.

Here, we focus rather narrowly on private consumption as our measure of welfare. Such a monetary measure of welfare is useful as a single indicator because it is likely to affect several dimensions of well-being, notably aspects of material well-being. As such monetary measures are arguably among the most comprehensive single measures of welfare. It is possible to try to construct composite indices of welfare that cover more dimensions of well-being, but it is often hard to quantify some non-material aspects of well-being and the weights used in aggregating different aspects are inevitably rather arbitrary. A preferable procedure is probably to look at monetary measures of welfare while at the same time looking at other single indicators that measure different dimensions of well-being such as education and health outcomes. These can be regarded as important aspects of well-being and provide additional information to the monetary statistics presented here. Although non-monetary dimensions of well-being are likely to be correlated with monetary measures of well-being, the correlation is far from perfect.

The consumption aggregate we use as our measure of monetary welfare covers a wide variety of food and non-food purchases by the households along with consumption of some home-produced items, notably food crops. One would expect a household's total consumption to be highly correlated with its income. There are three reasons for preferring consumption to income as a measure of monetary welfare. First, consumption may be a better measure of a household's long-term income than income in any one year. Annual incomes may fluctuate due to variations in the harvest or other temporary changes, but households are likely to use saving and borrowing try to smooth their consumption in the face of such transient changes in income. Secondly, in developing countries where most people work as smallholders or in informal enterprises, consumption may be more accurately measured by surveys than income is. Thirdly, it might be held that what people actually consume with their money affects their wellbeing more than what they simply earn.

A serious limitation with monetary measures of welfare is that they are typically observed only at the household, rather than individual, level. It is very hard to determine what each individual in the household consumes and conventionally surveys do not attempt to do this. Consequently, if households do not share consumption equally among their members, it is likely that monetary measures of well-being under-estimate inequality and poverty. Since households are of different sizes, it is common to look at household consumption per capita. However, household members of different ages and sexes have different needs. For example, the WHO estimates calorie requirements to vary with age and sex. We allow for this by looking at the number of "adult equivalents" in a household, where the adult equivalence scales are based partly on calorie requirements. Our reference person is an adult male aged between 18 to 30 years. For example, the WHO estimates that a one-year old boy requires 1,200 calories per day and while a man engaged in subsistence farming requires around 3,000 calories. Hence we treat a one-year old boy as being equivalent to 0.40 of an adult male. Our welfare measure is thus total household consumption divided by the total number of adult equivalents in the household.

8.2   *Setting the poverty line*

Given a monetary measure of welfare, we assess whether people are poor according to whether their level of welfare falls below the poverty line. Conceptually, the poverty line is the level of welfare that is regarded as the minimum people can enjoy without being regarded as poor. However, setting such a poverty line in practice is problematic and ultimately involves a large amount of judgement about what individuals need. Part of the problem is that it is impossible to draw a precise line that meaningfully distinguishes between people just on either side of the line. For example, if we use a "dollar a day" poverty line, it is untenable to argue that those existing on one dollar are significantly better off than those existing on 99 cents. This is not a fatal problem with poverty lines, as they can still provide useful information. It is an argument for not focussing too exclusively on particularly poverty

measures, but also looking more broadly at changes in welfare across the lower part of the income distribution.

More seriously, people are likely to have very different judgements about what are basic needs. As countries develop, norms about what is a reasonable standard of living are likely to be raised and so, in some sense, poverty in a particular country is likely to be relative, that is, defined relative to the average living standards prevailing in that country. However, measuring poverty using a relative poverty line can be misleading when trying to measure changes in wellbeing of poorer people. For example, if poor people's welfare increases, but at a slower rate than the welfare of others, then poverty defined relative to average living standards may rise even though the poor are in fact better off. In what follows, we fix the poverty line over time so that it does not vary with the average level of welfare in the country. Hence, on our measures, poverty will change if and only if the actual living standards of the poor change. As Uganda develops, there will be an argument for reviewing the poverty line to match changing views of what is regarded as acceptable minimum poverty levels. But, when measuring development in the short term, it is more sensible to fix the line.

When deciding what level to fix a poverty line at, a common procedure in developing countries is to anchor the line according to some basic needs and to food needs in particular. In developing countries such as Uganda, food accounts for around a half of all consumption. No one could disagree that food is an important need – being necessary for survival, for health and for activities of daily life. Moreover, it is possible to assess food requirements with reasonable objectivity. In particular, the calories required to perform various tasks can be estimated and there is a degree of consensus around the benchmarks for calorie requirements set by WHO (1985). It could be questioned why there is an exclusive concern with calories, rather than looking at other aspects of food consumption. Typically, however, people eating sufficient calories are also found to be meeting their protein requirements. Deficiencies in specific minerals and vitamins may remain, but these kinds of deprivation may require more targeted nutritional interventions and are not necessarily linked to general economic deprivation. When setting a poverty line, allowance is made for the kinds of foods people actually eat, which in turn reflect wider considerations than just their calorific value.

We work with a poverty line that reflects the cost of meeting calorie requirements given the typical diets of poor Ugandans, and an estimate of meeting non-food requirements. According to the principles set out by WHO (1985), a man (18-30 years) working in subsistence agriculture requires around 3000 calories per day. Consequently, we set our food poverty line at the cost of meeting that requirement. Women and children typically require fewer calories and this is taken into account by comparing household consumption per adult equivalent (rather than per capita) with the poverty line. Many combinations of foods ("food baskets") could meet the requirement of 3000 calories. We focus on the food basket of the poorest 50% of Ugandans, ranked by consumption per adult equivalent. We use data from the 1993/94 First Monitoring Survey to identify the mean quantities of different food items consumed by the poorest 50%. This calorific value of this basket was estimated and then the quantity of food in the basket was scaled up so that it provided exactly 3,000 calories per day. The cost of this food basket was then taken to be the food poverty line. It should be noted that this is a national food basket, although in practice people in different regions of the country tend to eat different staple foods. The use of a national food basket may be more appropriate if we wish to assess the capacity of people to obtain sufficient calories, although regional food baskets may be more appropriate if we wish to assess whether the sufficient calories are actually obtained.

Although we specify a food poverty line based on detailed itemisation of needs, such a procedure is very problematic when applied to non-food needs. Non-food expenditures are so

varied, the degree of subjectivity in specifying minimum requirements would make achieving consensus difficult. Instead, we follow the standard practice of simply making non-food requirements a mark-up on food requirements. Specifically, we follow Ravallion and Bidani (1994) in identifying non-food requirements as the non-food expenditure of those whose expenditure is just equal to the food poverty line. The rationale for this is that, since at this level of welfare the poor have sacrificed some of their need for calories, the non-food expenditures they have chosen to give priority to should also be regarded, as meeting essential needs. We allow different locations (Central urban, Northern rural etc) to have different non-food requirements. This allows for the fact that people in urban areas typically spend a higher share of their budget more on non-food items, even controlling for income for a variety of reasons (such as higher housing costs and greater transport costs in getting to work).

As a result of non-food requirements being allowed to vary with location, we do not use one single "all Uganda" national poverty line. However, averaging across Uganda, the poverty line(s) came to around $34 per capita per month in 1993/94 and hence were comparable the "$1 a day" poverty line sometimes used for international poverty comparisons by the World Bank.

8.3   Aggregation over individuals

Given a welfare measure (consumption per adult equivalent) and a poverty line, we can identify which Ugandans are poor. The final issue in measuring poverty is to aggregate this information to obtain a single poverty statistic for Uganda. This is an example of an "index number problem", in that we must reduce a vector – poverty status of millions of Ugandans – to a single scalar value.

We present the "Foster-Greer-Thorbecke" or "P-alpha" class of poverty indicators. These are defined generally as expressed in equation (3):

$$(3)\ P_{\alpha} \equiv \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\max[z - c_i; 0]}{z} \right\}^{\alpha}$$

where z= poverty line; and c_i =welfare.

Three variants of these indicators are presented, according to the value of α:

1) P_0, the poverty headcount, gives the percentage of Ugandans living below the poverty line (H=q/n). This measure is very intuitive and easy to popularise. However, it has a serious conceptual deficiency in that it is insensitive to changes in the welfare of people below the poverty line (this is termed violating the principle of monotonicity). It would be possible for the welfare of all the poor to be halved and as long as the non-poor were not affected, the poverty headcount would be unchanged.

2) P_1, the poverty gap indicator, measures how far the welfare of the poor lies below the poverty line. It is measured as: $P_1 = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\max[z - c_i, 0]}{z} \right\}$. Verbally, it can be thought of as showing the cost of eliminating poverty through perfectly targeted transfers to the poor, expressed as a fraction of the poverty line per Ugandan. (So if $P_1 = 0.1$, eliminating the poverty gap through perfect transfers would cost 10% of the poverty line per Ugandan.) In practice, it is impossible to perfectly target transfers (ie to give the poor (only) exactly enough money to raise their consumption to just above the poverty line). The advantage of the poverty gap measure over the headcount is that the poverty gap is sensitive to changes in the welfare of the poor. It has two disadvantages. Firstly, it is rather less intuitive and harder to publicises. Secondly, it is not sensitive to redistribution of welfare

among the poor. For example, if money was taken from the less poor to the extremely poor, we would tend to conclude that this reduced poverty but the P1 indicator would be unchanged (this is termed violating the principle of transfers).

3) P2, the squared poverty gap, $P_2 = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\max[z-c_i,0]}{z} \right\}^2$. This measure is sensitive to redistribution amongst the poor but is the least intuitive of the *P*-alpha measures.

In practice, the three *P*-alpha indicators often tend to move in a similar direction and so choosing between them is seldom required. A great advantage of the *P*-alpha class of indicators is that they are additively decomposable. For example, if we split the population into two groups (say urban and rural), then national poverty indicator is equal to the sum of the poverty indicators for the two groups, weighted by their population shares, that is,

$P = \sum_{j=1}^{j} \frac{n_j}{n} P_j$. It is also possible to conduct statistical testing using the indicators, with the formulae for their standard errors being given in Kakwani (1990).

8.4 Data transformation

UNHS-2 shared very similar consumption sections, with almost the same list of item codes and identical recall periods as UNHS I. Although, the former includes a few items not listed separately in the latter survey, these changes are minor and mainly reflect new areas of consumption such as mobile phones. Different recall periods were used to capture information on different sub-components of household expenditures. While a 7-day recall period was used for expenditure on food, beverages and tobacco, a 30-day recall period was used in the case of household consumption expenditure on non-durable goods and frequently purchased services. For the non-consumption expenditures and semi-durable and durable goods and services a 365-day recall period was used.

In the survey, all purchases by household members and items received free as gifts were valued and recorded as per the current prices. The items consumed out of home produce were valued at the current farm-gate/producer prices while rent for owner occupied houses was also imputed at current market prices. Food consumption includes food consumed from own production, purchases and free collection/gifts. We imputed house rent for some 163 households who had missing rent values via a hedonic model.

Expenditure data are collected on item by item basis. The expenditures were aggregated according to the recall period used and by broader sub-components of expenditures to a household level. Given the different recall periods used to collect data on household expenditures, some conversion factors were applied to change the data on a monthly basis. After which all the different sub-components of the expenditures were aggregated to derive the total consumption expenditures at household level. The consumption expenditures exclude non-consumption expenditures such as expenses on funeral, taxes and dividends etc. Although the administered to 9,711 households, the consumption aggregate was constructed for only 9,710 households. One household identified as hh=40430034803 was dropped because the food expenses recorded were for a ceremony, which took place during the survey period.

Although simply comparing nominal estimates of consumption expenditures with the consumer price index is useful to obtain a ball-park figure for real private consumption, we make two further adjustments for price effects when estimating poverty. Specifically, we re-value home consumption of food into market prices and we adjust for regional differences in food prices and finally accounting for household composition in terms of sex and age. The food price index is used in the main estimates to adjust for regional variations in food prices.

However, it is not used to adjust for inflation – i.e. it set to average 100 in each survey and only the CPI is used to adjust for inter-temporal variation in prices.

As a cautionary note, comparison of poverty trends in Uganda should take into account the changes in geographical coverage across survey rounds. As previously pointed out, some districts were not covered due to insurgencies at the time of the survey.

References

Appleton (2001a) "Changes in poverty in Uganda, 1992-1997", chapter in P. Collier and R. Reinnikka (eds.) "Firms, households and government in Uganda's recovery", World Bank: Washington DC

Appleton (2001b) "Poverty in Uganda, 1999/2000: Preliminary estimates from the UNHS", mimeo, School of Economics, University of Nottingham: Nottingham.

Datt, G. and M.Ravallion (1992) "Growth and redistribution components of changes in poverty measures: a decomposition with application to Brazil and India in the 1980s" *Journal of Development Economics* 38:275-295

Foster, J., Greer, J. and Thorbecke, E. (1984) "A Class of Decomposable Poverty Measures", *Econometrica*, 52: 761-6.

Kakwani, N., "Statistical inference in the measurement of poverty", *Review of Economics and Statistics*, 75(4):632-639.

Ravallion, Martin and Benu Bidani (1994) "How robust is a poverty line?" *World Bank Economic Review* 8(1): 75-102.

Uganda Bureau of Statistics (2003), *Administrative Report*, Entebbe.

Appendices

Appendix 1: How to obtain copies of the data and documentation