Household Survey Data Basics

Jann Lay

Kiel Institute for the World Economy

Overview

I. Background

II. Household surveys

- Design
- Content
- Quality
- Availability

I. Background

- Not new, household survey data has been collected for some decades (India: Regular surveys since the 1940s)
- However, until the 1990s:
  - Questionable quality
  - In many countries: changes in frequency of conducting surveys, changes in survey design etc.
  - No transparent access policies
  - Analysis of large survey datasets cumbersome and expensive
- In the 1990s, the collection, use, and analysis of household survey data change dramatically due to two factors

  (1) Technical progress

  (2) The Living Standards Measurement Study (LSMS) of the World Bank

(1) Technical progress

- Collection (Use of Laptop computers in the field)
- Availability (More timely availability)
- Analysis

(2) LSMS

- Original idea: datasets for poverty and distributional analysis comparable across countries
- In light of theoretical developments regarding household behaviour (Schultz and Becker): Not only measure welfare, but understand determinants of welfare (and household behaviour)
- Result: Comprehensive integrated surveys covering many important aspects of economic and domestic activity (today slight reversal towards emphasis on measurement)
- Since the mid-1980s a number of LSMS surveys have been carried out in many different countries (In Africa: Côte d'Ivoire, Ghana, Mauritania, Morocco, the Kagera region of Tanzania, South Africa)
- LSMS experiences with important impact on design, collection of surveys, and access policies (Best-practices)

II. Household surveys: Design, content, quality, and availability

Household Surveys: Design

- A typical household survey collects data on a national sample of households, randomly selected from a list of households (Census), although sometimes surveys are confined to particular regions
- Sample size varies – typical: in a population of 20 million individuals (4 million households) 50 000 individuals (10 000 households) in the sample
- Common: three-stage sample design
  (1) Defining strata (e.g. urban-rural, districts)
  (2) Selecting clusters or primary sampling units – PSUs (e.g. villages, blocks)
  (3) Selecting households

Ad (1) Defining strata

- Usually, surveys are required to generate statistics at least by geographical area
- Stratification e.g. by geographical areas splits a single survey into multiple surveys and guarantees in advance that there will be enough observations for each area for estimations
- Statistical advantage: Stratification reduces the variance of estimates by using prior information

Ad (2) Selecting clusters

- Clustering reduces costs
- Makes it worthwhile to collect village or community information
- Statistical disadvantage: Clustering increases the variance of estimates

- Weights and inflation factors:
  - Surveys can be stratified and clustered, and each household can still have the same probability to be selected into the sample
  - However, in reality probabilities of selection will differ due to
    - Higher costs of selecting certain households (e.g. those in remote areas)
    - Higher likelihood of refusing the interview of certain households
    - Differential selection probabilities (If you want to estimate mean income in rural areas, rich households should be overrepresented to enhance precision!)
  - When selection probabilities differ, we need to undo the sample design and weight the observations according to the selection probability
    - The higher the selection probability, the lower the weight given to the household, or
    - The lower the selection probability, the more households does one observation represent

Household Surveys: Content

- Data are gathered for all individuals in a household
- Definition of household difficult
  - Living arrangements not constant across countries and across time (E.g. polyandry, each wife runs own household)
  - Households function as production and consumption units, the members of these two units are not necessarily the same
  - Household may form only temporarily in response to external conditions
  - However, as every individual reports its affiliation to the household, the decision is with the analyst

- There are different types of surveys with different foci, e.g.
  - DHS (Demographic and Health Surveys)
  - Budget (Expenditure) surveys
  - Labour force (Income and employment) surveys
  - Informal Sector surveys
  - Agriculture surveys
  - LSMS-type surveys
- E.g. Uganda has a LSMS-type survey with changing focus modules (agricultural sector module in 1999/2000, employment module in 2002/03)

- Examples of demanded information in an LSMS-type survey (here Ugandan survey from 1999/2000):

(1) On the household level

- Dwelling characteristics (e.g. building material, hygienic facilities, piped water, access to firewood etc.)
- Expenditure
- Shocks experienced by the household
- Physical assets, loans

(2) On the individual level

- Sex, age, marital status
- Education, literacy, distance to school
- Health, health care for women, anthroprometics, fertility
- Migration
- Activity status, sector, occupation

- Specialised surveys have more detailed questions

- Getting practical:

(1) Example from the questionaire of the 1993 LSMS survey from South Africa

  - From http://www.worldbank.org/lsms/
  - Dataset available on the web (for free)
  - Here: an exempt from the section on dwelling characteristics

(2) How does the data look like in its ready-to-use format?

Section 2: Household Services

2.1 Housing

1. Type of dwelling?

Shack ..... -01
House/Part of a house ..... -02
Traditional dwelling (hut) ..... -03
Maisonette 04
Flat ..... -05
Hostel ..... -06
Outbuilding ..... -07
Combination of buildings ..... -08
Other (Specify) ..... -09
.....

2. What are the main materials used for the roof, wall and floor?
Circle the appropriate codes. (Not more than 3)

	2a. Roof	2b. Walls	2c. Floor Covering/Floor
Bricks	-01	-01	-01
Cement block	-02	-02	-02
Pre-fab	-03	-03	-03
Corrugated iron	-04	-04	-04
Wood	-05	-05	-05
Plastic	-06	-06	-06
Cardboard	-07	-07	-07
Mixture of mud and cement	-08	-08	-08
Wattle and daub	-09	-09	-09
Tile	-10	-10	-10
Carpet			-11
Linoleum			-12
Mud	-13	-13	-13
Thatching	-14	-14	-14
Asbestos	-15	-15	-15
Other (Specify)	-16	-16	-16

3. How many rooms does the dwelling have? (Exclude bathrooms, toilets & passages but include kitchens, lounges & dining rooms)

_____

4. How many rooms does the household occupy in this dwelling?

_____

5. Does the household own this dwelling?

Yes ..... -1
No ..... -2 → **Go to Question 9**

6. **If owned:** Does the household have a bond or loan on this dwelling?

Yes ..... -1
No ..... -2 → **Go to Question 8**

7. **If YES:** How much is still owed on the bond or loan?

R _____

8. If you sold this dwelling today, about how much do you think you could get for it? (i.e. how much would it cost to rebuild it today?)

R _____ → **Go to Section 2.2**

9a. Does the household have to pay rent to live here?

Yes ..... -1
No ..... -2 → **Go to Question 9c**

9b. **If YES:** How much rent is the household supposed to pay in a month?

R _____

9c. **If NO:** Suppose you did have to pay rent to live here, how much do you think you would have to pay in a month?

R _____

Stata Editor

Preserve | Restore | Sort | << | >> | Hide | Delete...

hhid[135] = 1013100003207

	hhid	hhweight	stratum	ea	per001	sex	age	edu
135	1013100003207	73	101	3	household head	male	43	do
136	1013100003207	73	101	3	spouse	female	34	Completed P.7
137	1013100003207	73	101	3	child	male	16	Completed P.3
138	1013100003207	73	101	3	child	female	10	Completed P.4
139	1013100003207	73	101	3	child	female	16	Completed P.7
140	1013100003207	73	101	3	child	female	14	Completed P.7
141	1013100003207	73	101	3	other relatives	female	4	No schooling and no f
142	1013100003207	73	101	3	other relatives	female	8	Completed P.2
143	1013100003207	73	101	3	other relatives	female	19	Completed S.2
144	1013100003207	73	101	3	other relatives	female	17	Completed S.2
145	1013100003208	73	101	3	household head	male	26	Completed P.4
146	1013100003208	73	101	3	spouse	female	21	Completed P.6
147	1013100003208	73	101	3	child	male	2	No schooling and no f
148	1013100003208	73	101	3	child	male	5	No schooling and no f
149	1013100014101	59	101	14	household head	male	46	Completed P.4
150	1013100014102	59	101	14	household head	female	68	No schooling and no f
151	1013100014102	59	101	14	other relatives	male	10	Some schooling but no
152	1013100014102	59	101	14	other relatives	male	9	Some schooling but no
153	1013100014102	59	101	14	other relatives	female	5	Some schooling but no
154	1013100014102	59	101	14	other relatives	female	9	Completed P.1
155	1013100014102	59	101	14	other relatives	male	14	Completed P.6
156	1013100014102	59	101	14	other relatives	male	17	Completed S.2
157	1013100014102	59	101	14	other relatives	female	20	Completed S.4
158	1013100014103	59	101	14	household head	male	26	Completed P.7
159	1013100014103	59	101	14	spouse	female	25	Completed P.6
160	1013100014103	59	101	14	child	female	0	No schooling and no f
161	1013100014103	59	101	14	child	male	6	No schooling and no f
162	1013100014103	59	101	14	child	female	3	No schooling and no f
163	1013100014103	59	101	14	other relatives	female	8	Completed P.2

Household Surveys: Quality

- Overall, the quality of surveys has improved and continues to improve
- problems remain, some of which are hard to tackle
- Two types of errors:

(1) Sampling errors

- Arise due to looking at a sample only, and not the entire population
- Decrease with sample size (though not proportionally)
- Can be reduced by adequate sample plan and accounted for by using weights (remember stratification and clusters!)

(2) Measurement (non-sampling) errors

- Random errors (are not a problem as they tend to cancel out with increasing sample size)
- In contrast, all the following errors are systematic
- Coverage errors (e.g. interviewers miss households), faulty frame
- Errors related to questionaire design (e.g. difficult questions) and data collection (e.g. interviewer bias)
- Respondent bias (e.g. sensitive topics , recall bias, incentives to understate or overstate)
- Processing mistakes

- More problems and possible sources of error arise when the data are analysed and derivative variables, for example household consumption, are calculated

- Calculating a consumption aggregate from raw data on
  - purchased goods
  - goods consumed from home produce
  - freely available goods

  (for each single good you have the quantity and the price for purchased goods)
- A flavour of the difficulties:
  - Which prices do you assign to home produce and goods available for free for some and not for others (e.g. firewood)?
  - What is the value of a TV, a sofa, a washing machine?
  - Serious underreporting of tobacco and alcohol consumption
  - The degree of disaggregation of goods consumed may influence the results (How much did you spend on food? vs. How much did you spend on rice, potatoes, matoke, pork meat, beef, fish, etc.?)

Household Surveys: Availability

- Availability and accessibility has improved dramatically in recent years
- Some LSMS surveys can even be downloaded from the World Bank's LSMS page
- The World Bank has launched the Africa Household Survey Databank at

  http://www4.worldbank.org/afr/poverty/databank/default.cfm

Region or Country	Survey Type	Sort by	Display
Africa	All Countries	All Types	Country,Year,Type	20 surveys

Found 406 surveys out of 406

Page 1 on 21 >> Next 20 survey(s)

Country	Year	Survey Name	Type	Info
Angola	1995	Inquerito Prioritario Sobre as Condicoes de Vida dos Domicilios	PS	****
Angola	1992	Inquerito sobre emprego e desemprego na cidade de Luanda	LAB	****
Angola	1990	Household Budget and Nutrition Survey	IES	***
Benin	2002	CWIQ-Social Fund Impact 2002 (Planning stage)	CWI	*
Benin	2001	Enquête Démographique et de Santé 2001	DHS	**
Benin	1996	Enquête Démographique et de Santé 1996	DHS	****
Benin	1996	Enquête sur les dépenses des ménages de Cotonou 1996 (UEMOA)	IES	****
Benin	1994	Etude sur les conditions de vie des ménages ruraux au Bénin (ECVR)	IS	*
Benin	1986	Enquête Budget-Consommation	IES	*
Botswana	2000	Multiple Indicators Cluster Survey	MIC	***
Botswana	1988	Family Health Survey 1988	DHS	***
Botswana	1985	Household Income and Expenditure Survey	IES	*
Burkina Faso	2003	Enquête Démographique et de Santé (Data processing in progress)	ND	**
Burkina Faso	2002	Etude sur les conditions de vie des ménages / CWIQ 2002 (Planning stage)	CWI	*
Burkina Faso	1998	Enquête démographique et de santé 1998/99	DHS	**