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 questionnaire 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)

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

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

6. If owned: Does the household have a bond or loan on this dwelling?
   - Yes ..................................... -1
   - No ..................................... -2

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 ____________________________

9a. Does the household have to pay rent to live here?
   - Yes ..................................... -1
   - No ..................................... -2

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 ____________________________
<table>
<thead>
<tr>
<th>hhid</th>
<th>hhweight</th>
<th>stratum</th>
<th>ea</th>
<th>per001</th>
<th>sex</th>
<th>age</th>
<th>edu</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>male</td>
<td>43</td>
<td>do</td>
</tr>
<tr>
<td>136</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>34</td>
<td>Completed P.7</td>
</tr>
<tr>
<td>137</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>male</td>
<td>16</td>
<td>Completed P.3</td>
</tr>
<tr>
<td>138</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>10</td>
<td>Completed P.4</td>
</tr>
<tr>
<td>139</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>16</td>
<td>Completed P.7</td>
</tr>
<tr>
<td>140</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>14</td>
<td>Completed P.7</td>
</tr>
<tr>
<td>141</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>4</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>142</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>8</td>
<td>Completed P.2</td>
</tr>
<tr>
<td>143</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>19</td>
<td>Completed S.2</td>
</tr>
<tr>
<td>144</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>17</td>
<td>Completed S.2</td>
</tr>
<tr>
<td>145</td>
<td>1013100003207</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>male</td>
<td>26</td>
<td>Completed P.4</td>
</tr>
<tr>
<td>146</td>
<td>1013100003208</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>female</td>
<td>21</td>
<td>Completed P.6</td>
</tr>
<tr>
<td>147</td>
<td>1013100003208</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>male</td>
<td>2</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>148</td>
<td>1013100003208</td>
<td>73</td>
<td>101</td>
<td>3</td>
<td>male</td>
<td>5</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>149</td>
<td>1013100014101</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>46</td>
<td>Completed P.1</td>
</tr>
<tr>
<td>150</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>68</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>151</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>10</td>
<td>Some schooling but no</td>
</tr>
<tr>
<td>152</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>9</td>
<td>Some schooling but no</td>
</tr>
<tr>
<td>153</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>5</td>
<td>Some schooling but no</td>
</tr>
<tr>
<td>154</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>9</td>
<td>Completed P.1</td>
</tr>
<tr>
<td>155</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>14</td>
<td>Completed P.6</td>
</tr>
<tr>
<td>156</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>17</td>
<td>Completed S.2</td>
</tr>
<tr>
<td>157</td>
<td>1013100014102</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>20</td>
<td>Completed S.4</td>
</tr>
<tr>
<td>158</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>26</td>
<td>Completed P.7</td>
</tr>
<tr>
<td>159</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>25</td>
<td>Completed P.6</td>
</tr>
<tr>
<td>160</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>0</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>161</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>male</td>
<td>6</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>162</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>3</td>
<td>No schooling and no f</td>
</tr>
<tr>
<td>163</td>
<td>1013100014103</td>
<td>59</td>
<td>101</td>
<td>14</td>
<td>female</td>
<td>8</td>
<td>Completed P.2</td>
</tr>
</tbody>
</table>
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 questionnaire 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
<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Survey Name</th>
<th>Type</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>1996</td>
<td>Inquérito Prioritário Sobre as Condições de Vida dos Domicílios</td>
<td>PS</td>
<td>****</td>
</tr>
<tr>
<td>Angola</td>
<td>1992</td>
<td>Inquérito sobre emprego e desemprego na cidade de Luanda</td>
<td>LAB</td>
<td>****</td>
</tr>
<tr>
<td>Angola</td>
<td>1990</td>
<td>Household Budget and Nutrition Survey</td>
<td>IES</td>
<td>***</td>
</tr>
<tr>
<td>Benin</td>
<td>2002</td>
<td>CWIG-Social Fund Impact 2002 (Planning stage)</td>
<td>CWI</td>
<td>*</td>
</tr>
<tr>
<td>Benin</td>
<td>2001</td>
<td>Enquête Démographique et de Santé 2001</td>
<td>DHS</td>
<td>**</td>
</tr>
<tr>
<td>Benin</td>
<td>1996</td>
<td>Enquête Démographique et de Santé 1996</td>
<td>DHS</td>
<td>****</td>
</tr>
<tr>
<td>Benin</td>
<td>1996</td>
<td>Enquête sur les dépenses des ménages de Cotonou 1996 (UEMCA)</td>
<td>IES</td>
<td>****</td>
</tr>
<tr>
<td>Benin</td>
<td>1994</td>
<td>Etude sur les conditions de vie des ménages ruraux au Bénin (ECVR)</td>
<td>IS</td>
<td>*</td>
</tr>
<tr>
<td>Benin</td>
<td>1986</td>
<td>Enquête Budget-Consortinun</td>
<td>IES</td>
<td>*</td>
</tr>
<tr>
<td>Botswana</td>
<td>2000</td>
<td>Multiple Indicators Cluster Survey</td>
<td>MIC</td>
<td>***</td>
</tr>
<tr>
<td>Botswana</td>
<td>1998</td>
<td>Family Health Survey 1998</td>
<td>DHS</td>
<td>***</td>
</tr>
<tr>
<td>Botswana</td>
<td>1985</td>
<td>Household Income and Expenditure Survey</td>
<td>IES</td>
<td>*</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2003</td>
<td>Enquête Démographique et de Santé (Data processing in progress)</td>
<td>ND</td>
<td>**</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2002</td>
<td>Etude sur les conditions de vie des ménages / CWIG 2002 (Planning stage)</td>
<td>CWI</td>
<td>*</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>1998</td>
<td>Enquête démographique et de santé 1998/99</td>
<td>DHS</td>
<td>**</td>
</tr>
</tbody>
</table>