Introduction to the Timor Leste Living Standards Survey (TLSLS)

Training Workshop on the Commitment to Equity Methodology
CEQ Institute, Asian Development Bank, and The Ministry of Finance
Dili
June, 2017
About Household Surveys

• How GDE carries them out
  • draw a random sample of households
    • stratified by populations of interest
    • clustered, for economy
  • ask them a lot of questions
    • see the questionnaire

• Why GDE carries them out
  • useful to establish consumption patterns to establish weights for the consumer price index
  • useful to estimate (or describe) the distribution of income and/or consumption in Timor Leste
    • poverty
    • inequality
  • keeps researchers busy
About TLSLS

• How the data are structured
  • Many files (look at subdirectories)
  • More or less follow the structure of the questionnaire
  • Some are “household-level” data
    • one record per household
  • Some are individual-level, and some are community-level

• Skip patterns and missing values
  • look at section 4, part A, for example
  • Stata does not behave well with missing values, so be careful!
About Sampling and Sampling Weights

• Not every observation in the GLSS is equally valuable
  • urban areas are “over-sampled” because it is easy to interview there
  • there could be other over-sampled groups
    • for example, a small ethnic minority
    • Oecussi

• Because it is more likely that an urban household is chosen for the sample, it gets less weight than a rural one

• Intuition: the rural households selected must “represent” more households than the urban households selected

• A simple example
## Simple Example of Sampling Weights

<table>
<thead>
<tr>
<th>Person</th>
<th>Area</th>
<th>Income</th>
<th>Sampled?</th>
<th>Sampled Income</th>
<th>Probability of Selection</th>
<th>Sampling Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Urban</td>
<td>100</td>
<td>Yes</td>
<td>100</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Urban</td>
<td>300</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Urban</td>
<td>200</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Urban</td>
<td>400</td>
<td>Yes</td>
<td>400</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Rural</td>
<td>50</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Rural</td>
<td>100</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Rural</td>
<td>20</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Rural</td>
<td>30</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Rural</td>
<td>40</td>
<td>Yes</td>
<td>40</td>
<td>0.33</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Rural</td>
<td>60</td>
<td>Yes</td>
<td>60</td>
<td>0.33</td>
<td>3</td>
</tr>
</tbody>
</table>

### True Average of Income

<table>
<thead>
<tr>
<th>Person</th>
<th>True Average of Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average:</td>
<td>130</td>
</tr>
</tbody>
</table>

### Unweighted Estimate, Average Income

<table>
<thead>
<tr>
<th>Person</th>
<th>Unweighted Estimate, Average Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average:</td>
<td>150</td>
</tr>
</tbody>
</table>

### Weighted Estimate

<table>
<thead>
<tr>
<th>Person</th>
<th>Weighted Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average:</td>
<td>130</td>
</tr>
</tbody>
</table>
About Clustering and Stratification

• Random sampling of households vs. random sampling of clusters of households
  • see following graphics
Randomly drawn households
Randomly drawn *clusters* of households (census enumeration areas)
About Clustering and Stratification

- Clustering tends to increase the standard errors of things we estimate with survey data
  - households near each other are more similar to each other than they are to other households
  - reduces the variation in the sample

- Stratification can reduce the standard errors, but usually does not in TLSLS-type surveys

- Stata has very easy-to-use commands to take these things into account
What Does TLSLS Have for CEQ?

Market Income

Contributory Pensions

Market Income plus Pensions

Direct Transfers

Gross Income

Direct Taxes

Net Market Income

Non-Taxable Income

Taxable Income

Direct Transfers

Gross Income

Direct Taxes

Disposable Income

Indirect Subsidies

Consumable Income

Indirect Taxes

Final Income

In-Kind Transfers (Education, Health)

Copayments, User Fees
How Do We Go About the Analysis?

• We must dig up all the information we can about these income concepts and their components from TLSLS
• We must clean it up and save it in one dataset
• We can then apply pre-programmed commands to do much of the statistical work
• We must cross-check our results against administrative data whenever possible