Lab 2: Estimating Poverty Lines from Household Surveys
Short Course on Poverty & Development for Nordic Ph.D. Students
University of Copenhagen
June 13-23, 2000

In this lab you are provided a sample Living Standards Measurement Survey-type (LSMS) dataset with which to estimate a poverty line using the cost of basic needs (CBN) approach. Given the limited amount of time, we will not use the food energy intake (FEI) method. This does not mean that the FEI method is any less valid. But because the CBN method is less computationally straightforward than the far easier FEI method, we take this time to elucidate the mechanics behind it.

Further, so that we do not get bogged down in details, and so that you have exercises with which you can experiment, some major simplifying assumptions are made in the sample Stata program provided. These include: (a) a national food basket, (b) national food prices, and (c) national costs for non-food consumption. In other words, we are going to estimate a national poverty line. These assumptions are made to simplify the program so that you get the basic idea of how to estimate a poverty line.

Stata notes:

1. This sample program is written as a batch file using a basic text editor (e.g. Notepad). After you write it on your own, open Stata and change the directory by typing (the directory might be different)

```
cd c:\copen\lab2\
```

Then submit the program (saved as c:\copen\lab2\povline.do) by typing

```
do povline
```

2. This program involves the use of Stata's programming language to run do-loops through the food items. See the separate sheet on basics of programming in Stata. As you will see in the later labs, we will use programming extensively (and even create our own commands that we can call). To take full advantage of Stata, it's worth taking the time to understand how to program.

3. This program also uses Stata's matrix and scalars. If you have any experience using Gauss, MATS, or any other matrix programming language, this will be very familiar to you. The basic idea is that you can define a matrix (vectors are defined as matrices by Stata) or a scalar and Stata will keep it in memory so that you can use it to modify data or to print out results.

Basic Steps in Estimating a Poverty Line using the CBN Method:

Here we follow the procedure of Ravallion & Bidani (1994)

1. *Choose the reference group*. We identify the poorest 50 percent of the population as the reference group.
2. *Determine the reference food basket*. We use the food consumption behavior of the reference group to determine the quantities of basic food items that will make up the reference food basket. In this case, the basket is made up of the mean consumption levels

(purchased, auto-consumption, and in-kind payments) of 31 food items by the poorest 50 percent of the population (in per capita terms). Note: This is estimated to be a national reference.

3. *Estimate total calories for the reference food basket*. With information on the caloric content of each of the 31 food items, we estimate the total calories consumed by an individual who consumes this average basket.
4. *Scale the consumption levels by a constant so that the reference basket provides 2,400 calories per day*. The minimum level of calorie consumption is arbitrarily chosen to be 2,400 per day. The reference food basket estimated in (3) is unlikely to sum to this amount, so we scale the consumption levels up or down by a constant (so the consumption patterns do not change) to attain 2,400 calories.
5. *Cost the reference food basket to get a food poverty line*. Using price information (in this case from a community questionnaire, but this information can come from estimated median unit prices), we determine the cost of consuming the reference basket of 31 food items. Again note, we estimate a national food poverty line given "national" prices of the individual food items.
6. *Estimate the national poverty line*. We follow Ravallion & Bidani in estimating the non-food component of the poverty line by examining the consumption behavior of those households who can just afford the reference food basket. The non-food share of total expenditures is estimated by regressing the food share (s) of each household i on a constant and the log of the ratio of consumption expenditures to the food poverty line:

$$s_i = \alpha + \beta \log\left(\frac{y_i}{z^f}\right) + \varepsilon_i$$

For those whose expenditures are just equal to the food poverty line ($y_i = z^f$), the food share is α, and consequently the non-food share of expenditures is $(1-\alpha)$. Thus the poverty line is

$$z = z^f (2 - \alpha)$$

A sample program is attached below. In this lab, you can write it for yourself, experimenting along the way by looking at the data and what you get after each set of command. You can also experiment by estimating regional poverty lines by allowing for different regional patterns of food consumption, and by using `foodpcl2.dta` to estimate regional poverty lines.

```
version 6.0
clear
set matsize 350
set more off
set memory 5
#delimit ;

log using c:\copen\lab2\povline.log, replace;

************************************************

  c:\copen\lab2\povline.do

  Stata program to construct a poverty line using
  the cost of basic need approach for a sample
  dataset provided.

  The steps include:

    1. Choosing the reference group (poorest 50%)
    2. Determining the reference food basket
    3. Estimating the calories in the ref. basket
    4. Scaling the basket to get 2,400 calories
       (2,400 is chosen arbitrarily)
    5. Cost the basket to get the food poverty line
    6. Following Ravallion & Bidani, estimate non-food
       component as the non-food share of households who
       can just afford the basic food bundle. (See
       Ravallion & Bidani, 1994, p. 88)

  This program estimates a national poverty line.
  The lab exercise is to estimate regional poverty
  lines (Dataset to use is foodpcl2.dta).

************************************************;

* First define the minimum calorie consumption as 2,400 cal/day
  and convert it to an annual rate;

scalar mincal = 2400 * 365;

*--------------------------------------*
*--- 1. Choose the reference basket ---*
*--------------------------------------*;

use c:\copen\data\expend, clear;

gen expendpc = hhexpend / hhsize;
sum expendpc [aw=hhsize], detail;
scalar med = r(p50);

 ** Keep id codes of poorest 50% **;

keep if expendpc < med;
keep hid hhsize;
sort hid;
save c:\temp\temp, replace;

*----------------------------------------------*
*--- 2. Determine the reference food basket ---*
*----------------------------------------------*;

 ** merge household id (hid) of poorest
```

```
    50% with food consumption data     **;

use c:\copen\data\foodcon, clear;
sort hid;
merge hid using c:\temp\temp;
tab _merge;
drop if _merge==1;  * Keep only for the reference group *;
drop _merge;
replace quant = quant / hhsize; * Put cons in per capita terms *;

 ** Note: Stata only "replaces" a variable if an it is altered.
         So, if the hhsize is 1, then quant will not be altered;

 ***************************************************
 ** Program to generate variables for each food item **
 ** by looping through the item codes.               **;
 ** Then it estimates the means of each food item    **
 ** and saves this information as a dataset          **
 ***************************************************

capture program drop qtyprog;
 program define qtyprog;
 display "This will take some time...";

  qui sum item;
  local min = r(min);
  local max = r(max);
  local i = `min';
  while `i' <= `max' {;
   qui gen quant`i' = 0;
   qui replace quant`i' = quant if item == `i';
   local i = `i' + 1;
                      };

 ** Collapse the dataset so that there is only
    one observation per household.  Note, we use
    "max" so that hhsize (our weight variable) is
    unaltered;

 sort hid;
 collapse (max) hhsize quant`min'-quant`max', by(hid);
 display "";
 display "Means of the food items";
 display "";
 sum [aw=hhsize];

 ** Now estimate sample means for each individual
    food item, by looping through the items again.
    Note: We couldn't do this earlier because the
    data had yet to be collapsed
    NOTE FURTHER: This is where we could relax
    our assumption about a common national food
    basket by estimating the means over regions;

  local i = `min';
  while `i' <= `max' {;
   qui sum quant`i' [aw=hhsize];
    local mu = r(mean);
   if `i' == `min' {;
    matrix foodmu = ( `i',`mu' );
    matrix rownames foodmu = `i';
                   };
   if `i' ~= `min' {;
```

```
    matrix foodm = ( `i',`mu' );
    matrix rownames foodm = `i';
    matrix foodmu = foodmu \ foodm;
                  };
   local i = `i' + 1;
                    };
 matrix colnames foodmu = item mean;
 drop _all;  * Drop existing dataset, but not matrices;
 svmat foodmu, name(col);
 end;
 ************************************************;
qtyprog;

* This is our reference food basket *;
list;
compress;
sort item;

*-----------------------------------------------*
*--- 3. Estimate calories for this          ---*
*---    reference food basket               ---*
*-----------------------------------------------*;

merge item using c:\copen\data\foodpcl.dta;  * A small "L", not a "1";
tab _m;
drop _m;

  ** Sum all of the calories for the reference basket **;
preserve;
gen meancal = mean * calorie;
collapse (sum) meancal;
sum;

*---------------------------------------*
*--- 4. Scale the calories to 2,400 ---*
*---    for the reference basket    ---*
*---------------------------------------*;

display (mincal - r(mean))/365;
scalar cvfact = mincal/r(mean);
display "Scaling factor is " cvfact;
restore;
gen refmean = mean * cvfact;

*---------------------------------------*
*--- 5. Cost the reference basket   ---*
*---    to get the food poverty line---*
*---------------------------------------*;

gen cost = refmean * price;
preserve;
collapse (sum) refmean;
sum;
scalar zfood = r(mean);
display zfood;
restore;

*---------------------------------------*
*--- 6. Estimate national poverty   ---*
*---    line following Ravallion &  ---*
*---    Bidani (1994)               ---*
*---------------------------------------*;
```

```
use c:\copen\data\expend, clear;
gen expendpc = hhexpend / hhsize;
gen logyz = log(expendpc/zfood);
regress foodshr logyz [aw=hhsize];
scalar z = zfood*(2-_b[_cons]);
display zfood;
display z;
log close;
```