

# Econometric Game 2014

## Welfare analysis with incomplete expenditure data

### **Background.**

Every three years, Indonesia fields a national household survey (repeated cross section) that collects data on a wealth on household characteristics and expenditure. The survey is called the Susenas ( see <http://www.rand.org/labor/bps/susenas.html> ). The expenditure data form the basis of welfare statistics, including those on poverty. Expenditure data however are costly to collect. The full expenditure module requires households to answer around 600 questions on their expenditure, a time consuming task. Your assignment is to investigate whether proxies of wealth, which are easier to collect, are good predictors for the welfare statistics the Government is interested in tracking.

The data are from the Central Bureau of Statistics of Indonesia. They are for sale and cannot be distributed freely. To comply with this rule, the data made available for this game consist of an 80 percent random sample of the original data. You can use them for the game only.

## Case A

The full expenditure questionnaire is implemented for a random sub-sample of around 60,000 households. You have received the data for 1996,1999, and 2002 in one large file, where the expenditure data for 2002 have been discarded. The expenditure data are nominal figures, and thus have not been corrected for inflation. The data also contain several other household characteristics which are available for all years.

The poverty lines are in table 1. A household is deemed poor if their monthly per capita household expenditure – as measured by the full (module) questionnaire – is below the poverty line. These poverty lines can also be used as price deflators.

**Table 1 Poverty lines for Indonesia (Rupiah per person per month)**

year	Urban	rural
1996	42,032	31,366
1999	92,409	74,272
2002	130,499	96,512

Source: Indonesian bureau of statistics, c.f. [http://www.bps.go.id/tab\\_sub/view.php?kat=1&tabel=1&daftar=1&id\\_subyek=23&notab=7](http://www.bps.go.id/tab_sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=23&notab=7)

Your assignment is to complete the table 2 and table 3 with the best possible prediction based on the data you have received.

**Table 2 Welfare statistics for Indonesia in1996,1999,2002 in 1996 prices (standard deviations between parenthesis)**

	1996	1999	2002
<b>Average per capita expenditure</b>	71135	143590	
	(69929)	(291766)	
<b>Poverty head count ratio</b>	0.182	0.228	
	(0.386)	(0.419)	
<b>Poverty gap ratio</b>	0.034	0.042	
	(0.091)	(0.100)	

Table 3 Head count ratio by province in 1996 and 1999.

Province Code	Head count ratio 1996	Head count ratio 1999	Head count ratio 2002
12	0.135	0.179	
13	0.056	0.078	
14	0.047	0.076	
15	0.101	0.223	
16	0.141	0.211	
17	0.146	0.176	
18	0.270	0.324	
31	0.008	0.021	
32	0.116	0.185	
33	0.251	0.307	
34	0.184	0.233	
35	0.261	0.301	
51	0.069	0.055	
52	0.375	0.333	
53	0.445	0.573	
61	0.156	0.175	
62	0.050	0.050	
63	0.087	0.137	
64	0.042	0.123	
71	0.223	0.173	
72	0.194	0.282	
73	0.265	0.241	
74	0.291	0.306	

The poverty head count ratio and the poverty gap ratio are defined by

$$P_{\alpha} = \frac{1}{n} \sum_{y_i < p} \left( \frac{z - y_i}{z} \right)^{\alpha}$$

Where  $\alpha=0$  for the head count ratio and  $\alpha=1$  for the poverty gap ratio, and  $z$  is the poverty line. The data in the table above have been prepared with Stata. The program is copied in Appendix A.

The Module sample, which you have received, is a stratified random sample. The variable *modweight* is an expansion factor for the household. Thus, for each year;

$$\text{Population size} = \sum_{i=1}^M \text{hhmemb}_i * \text{modweight}_i$$

Where *M* is the number of households in the Module sample.

## Appendix A

```
* directories

global datadir "C:\Users\menno\Desktop"

use "$datadir\DataNewVarsCodesApril13.dta", clear

* these provinces do not have data for 2002
drop if prov==11 | prov==81 | prov==82 | prov ==54

* create an 80 percent random sample

set seed 100

gene uu=runiform()

drop if uu > .8

gene korweighti=korweight*hhmemb
gene modweighti=modweight*hhmemb

save temp,replace

collapse (sum) korweighti modweighti, by (year)

list

* Note. This shows that the weights add up to the population total

use temp,clear

save case,replace

keep if mod==1

egen cons=rsum( exf1-exf15 exnf1- exnf8)

gene picons=cons/ hhmemb

gene PL=.

replace PL=42032 if year==1996 & urru==1
replace PL=31366 if year==1996 & urru==2
replace PL=92409 if year==1999 & urru==1
replace PL=74272 if year==1999 & urru==2
replace PL=130499 if year==2002 & urru==1
replace PL=96512 if year==2002 & urru==2

gene poor=picons<PL
```

```
gene gap=(PL-pccons)/PL
```

```
replace gap=0 if poor==0
```

```
sort year
```

```
save temp,replace
```

```
by year: sum pccons poor gap [ aweight=modweighti]
```