



## USER MANUAL

**DASP version 3.02**

### **DASP: Distributive Analysis Stata Package**

By

**Abdelkrim Araar,  
Jean-Yves Duclos**

***Université Laval, PEP and World Bank***

**December 2021**

This work was carried out with financial and scientific support from the Partnership for Economic Policy (PEP), which is funded by the Government of Canada through the International Development Research Centre (IDRC) and Global Affairs Canada (GAC), and the William & Flora Hewlett Foundation. We are also grateful to the World Bank and UNDP that provided funds to develop the first version of DASP.

# Table of contents

Table of contents .....	2
List of Figures .....	5
1 Introduction .....	7
2 DASP and Stata versions.....	8
3 Installing and updating the <i>DASP</i> package.....	8
3.1    Installing DASP .....	8
3.2    Adding the DASP submenu to Stata's main menu .....	9
4 DASP and data files.....	10
5 Main variables for distributive analysis.....	11
6 How can DASP commands be invoked? .....	12
7 How can help be accessed for a given <i>DASP</i> module? .....	13
8 Applications and files in <i>DASP</i> .....	14
9 Basic Notation .....	16
10 DASP and poverty indices.....	16
10.1    Poverty indices (ipov) .....	16
10.2    Differences between poverty indices (dipov) .....	18
10.3    DASP and multidimensional poverty indices (imdpov) .....	18
10.4    Multiple overlapping deprivation analysis (MODA) indices .....	21
11 DASP, poverty and targeting policies.....	26
11.1    Poverty and targeting by population groups.....	26
11.2    Poverty and targeting by population group with a fixed budget (ogtpr) .....	27
11.3    Bi-dimensional poverty and targeting by population group (itargetg2d) .....	30
11.4    Poverty and targeting by income components .....	32
12 Marginal poverty impacts and poverty elasticities .....	33
12.1    FGT elasticity's with respect to the average income growth (efgtgr). .....	33
12.2    FGT elasticities with respect to average income growth with different approaches (efgtgro). 34	34
12.3    FGT elasticity with respect to Gini inequality (efgtineq).....	34
12.4    FGT elasticity with respect to Gini-inequality with different approaches (efgtine).....	36
12.5    FGT elasticities with respect to within/between group components of inequality (efgtg) .....	36
12.6    FGT elasticities with respect to within/between income components of inequality (efgtc) .....	38
13 DASP and inequality indices.....	39
13.1    Inequality indices (ineq) .....	39
13.1.1    Gini and concentration indices .....	40
13.1.2    Generalised entropy index.....	40
13.1.3    Atkinson index .....	41
13.1.4    Coefficient of variation index.....	42
13.1.5    Quantile/share ratio indices of inequality .....	42
13.2    Differences between inequality indices (dineq).....	42

13.3	The Araar (2009) multidimensional inequality index.....	43
14	DASP and polarisation indices.....	43
14.1	Polarisation indices (ipola).....	43
14.1.1	The Duclos, Esteban and Ray (2004) polarisation index.....	44
14.1.2	The Foster and Wolfson (1992) polarisation index .....	44
14.1.3	The generalised Esteban, Gradian and Ray (1999) polarisation index.....	45
14.1.4	The Inaki (2008) polarisation index.....	46
14.2	Differences between polarisation indices (dipola).....	48
15	DASP and decompositions.....	48
15.1	FGT Poverty: decomposition by population subgroups (dfgtg).....	48
15.2	FGT Poverty: decomposition by income components using the Shapley value (dfgts).....	49
15.3	Alkire and Foster (2011) MD index of poverty: decomposition by population subgroups (dmdafg).....	52
15.4	Alkire and Foster (2011: decomposition by dimensions using the Shapley value (dmdafs)....	52
15.5	Decomposition of the variation in FGT indices into growth and redistribution components (dfgtgr)53	
15.6	Decomposition of change in FGT poverty by poverty and population group components – sectoral decomposition- (dfgtg2d).....	54
15.7	Decomposition of FGT poverty by transient and chronic poverty components (dtcpov).....	56
15.8	Inequality: decomposition by income sources (diginis) .....	58
15.9	Regression-based decomposition of inequality by income sources.....	59
15.10	Gini index: decomposition by population subgroups (diginig).....	64
15.11	Generalised entropy indices of inequality: decomposition by population subgroups (dentropyg).....	65
15.12	Polarisation: decomposition of the DER index by population groups (dpolag).....	65
15.13	Polarisation: decomposition of the DER index by income sources (dpolas).....	66
16	DASP and curves.....	67
16.1	FGT CURVES (cfgt).....	67
16.2	Difference between FGT CURVES with confidence interval (dicfgt).....	68
16.3	Lorenz and concentration CURVES (clorenz).....	68
16.4	Differences between Lorenz/concentration curves with confidence interval (diclorenz).....	69
16.5	Poverty curves (cpoverty) .....	70
16.6	Consumption dominance curves (cdomc) .....	70
16.7	Difference/Ratio between consumption dominance curves (cdomc2d) .....	71
17	DASP and the progressivity.....	73
17.1	DASP and the progressivity indices .....	73
17.2	DASP and the progressivity curves .....	73
17.2.1	Checking the progressivity of taxes or transfers .....	73
17.2.2	Checking the progressivity of transfer vs tax .....	73
18	Dominance .....	74
18.1	Poverty dominance (dompov) .....	74
18.2	Inequality dominance (domineq) .....	74
18.3	DASP and bi-dimensional poverty dominance (dombdpov).....	75
19	Distributive tools.....	76
19.1	Quantile curves (c_quantile) .....	76

19.2	Income share and cumulative income share by group quantiles (quinsh) .....	76
19.3	Density curves (cdensity) .....	76
19.4	Nonparametric regression curves (cnpe).....	78
19.4.1	Nadaraya-Watson approach.....	78
19.4.2	Local linear approach .....	78
19.5	DASP and joint density functions.....	79
19.6	DASP and joint distribution functions.....	79
20	<i>DASP and pro-poor growth</i> .....	80
20.1	DASP and pro-poor indices.....	80
20.2	DASP and pro-poor curves.....	81
20.2.1	Growth incidence curve (gicur).....	81
20.2.2	Primal pro-poor curves .....	81
20.2.3	Dual pro-poor curves.....	81
21	<i>DASP and Benefit Incidence Analysis</i> .....	83
21.1	Benefit incidence analysis.....	83
21.2	Marginal benefit incidence analysis (imbi) .....	87
21.2.1	Derivative of the linear locally estimator approach.....	87
21.2.2	Ajwad and Quentin (2001) approach .....	88
22	Disaggregating grouped data .....	90
23	Appendices .....	95
23.1	Appendix A: illustrative household surveys.....	95
23.1.1	The 1994 Burkina Faso survey of household expenditures (bkf94l.dta) .....	95
23.1.2	The 1998 Burkina Faso survey of household expenditures (bkf98l.dta) .....	95
23.1.3	Canadian Survey of Consumer Finance (a subsample of 1000 observations – can6.dta)96	96
23.1.4	Peru LSMS survey 1994 (A sample of 3623 household observations - PEREDE94l.dta). 96	96
23.1.5	Peru LSMS survey 1994 (A sample of 3623 household observations – PERU_A_I.dta) ...97	97
23.1.6	The 1995 Colombia DHS survey (columbial.dta).....	97
23.1.7	The 1996 Dominican Republic DHS survey (Dominican_republic1996l.dta) .....	97
23.2	Appendix B: labelling variables and values.....	98
23.3	Appendix C: setting the sampling design .....	99
	Examples and exercises.....	101
23.4	Estimation of FGT poverty indices .....	101
23.5	Estimating differences between FGT indices .....	106
23.6	Estimating multidimensional poverty indices .....	110
23.7	Estimating FGT curves.....	113
23.8	Estimating differences between FGT curves with confidence intervals .....	120
23.9	Testing poverty dominance and estimating critical values.....	123
23.10	Decomposing FGT indices .....	124
23.11	Estimating Lorenz and concentration curves .....	126
23.12	Estimating Gini and concentration curves.....	131
23.13	Using basic distributive tools .....	134
23.14	Plotting the joint density and joint distribution function.....	140
23.15	Testing the bi-dimensional poverty dominance .....	143
23.16	Testing for pro-poorness of growth in Mexico .....	147
23.17	Benefit incidence analysis of public spending on education in Peru (1994) .....	153

## List of Figures

Figure 1: Ouput of <i>net describe dasp</i> .....	9
Figure 2: <i>DASP</i> submenu .....	10
Figure 3: Using <i>DASP</i> with a command window.....	12
Figure 4: Accessing help on <i>DASP</i> .....	13
Figure 5: Estimating FGT poverty with one distribution.....	14
Figure 6: Estimating FGT poverty with two distributions.....	15
Figure 7: The IMODA dialog box.....	23
Figure 8: Poverty and the targeting by population groups.....	27
Figure 9: Poverty and the targeting by population groups in case of fixed budget.....	28
Figure 10: Bidimensional poverty and the targeting by population groups .....	30
Figure 11: The EFGTG dialog box.....	37
Figure 12: The EFGTC dialog box.....	38
Figure 13: The IPOLA dialog box.....	47
Figure 14: Decomposition of the FGT index by groups.....	49
Figure 15: Decomposition of the FGT index by income components .....	51
Figure 16: Sectoral decomposition of FGT .....	55
Figure 17: Decomposition of poverty into transient and chronic components .....	57
Figure 18: Decomposition of the Gini index by income sources (Shapley approach) .....	59
Figure 19: FGT curves dialog box .....	67
Figure 20: Lorenz and concentration curves .....	69
Figure 21: Consumption dominance curves dialog box .....	71
Figure 22: IMBI dialog box .....	89
Figure 23: ungroup dialog box .....	93
Figure 24: Survey data settings .....	99
Figure 25: Setting sampling weights.....	100
Figure 26: Estimating FGT indices .....	103
Figure 27: Estimating FGT indices with relative poverty lines .....	104
Figure 28: FGT indices differentiated by gender .....	105
Figure 29: Estimating differences between FGT indices .....	107
Figure 30: Estimating differences in FGT indices .....	108
Figure 31: FGT differences across years by gender and zone .....	109
Figure 32: Estimating multidimensional poverty indices (A) .....	111
Figure 33: Estimating multidimensional poverty indices (B).....	112
Figure 34: Drawing FGT curves .....	114
Figure 35: Editing FGT curves .....	114
Figure 36: Graph of FGT curves.....	115
Figure 37: FGT curves by zone .....	116
Figure 38: Graph of FGT curves by zone.....	117

Figure 39: Differences of FGT curves .....	118
Figure 40: Listing coordinates .....	118
Figure 41: Differences between FGT curves.....	119
Figure 42: Differences between FGT curves.....	120
Figure 43: Drawing the difference between FGT curves with confidence interval .....	121
Figure 44: Difference between FGT curves with confidence interval ( $\alpha = 0$ ).....	122
Figure 45: Difference between FGT curves with confidence interval ( $\alpha = 1$ ).....	122
Figure 46: Testing for poverty dominance .....	123
Figure 47: Decomposing FGT indices by groups.....	125
Figure 48: Lorenz and concentration curves .....	127
Figure 49: Lorenz curves.....	128
Figure 50: Drawing concentration curves.....	128
Figure 51: Lorenz and concentration curves.....	129
Figure 52 :Drawing Lorenz curves.....	130
Figure 53: Lorenz curves .....	130
Figure 54: Estimating Gini and concentration indices .....	132
Figure 55: Estimating concentration indices .....	133
Figure 56: Estimating differences in Gini and concentration indices .....	134
Figure 57: Drawing densities.....	135
Figure 58: Density curves.....	136
Figure 59: Drawing quantile curves.....	137
Figure 60: Quantile curves.....	137
Figure 61: Drawing non-parametric regression curves.....	138
Figure 62: Non-parametric regression curves .....	138
Figure 63: Drawing derivatives of non-parametric regression curves.....	139
Figure 64: Derivatives of non-parametric regression curves .....	139
Figure 65: Plotting joint density function.....	141
Figure 66: Plotting joint distribution function .....	142
Figure 67: Testing for bi-dimensional poverty dominance .....	144
Figure 68: Drawing the growth incidence curve.....	148
Figure 69: Testing the pro-poor growth (primal approach).....	149
Figure 70: Testing the pro-poor growth (dual approach) – A .....	150
Figure 71: Testing the pro-poor growth (dual approach) – B .....	151
Figure 72: Benefit incidence analysis .....	153
Figure 73: Benefit Incidence Analysis (unit cost approach).....	155

# 1 Introduction

The Stata software has become a very popular tool to transform and process data. It comes with a large number of basic data management modules that are highly efficient for transformation of large datasets. The flexibility of Stata also enables programmers to provide specialised ".ado" routines to add to the power of the software. This is indeed how *DASP* interacts with Stata. *DASP*, which stands for *Distributive Analysis Stata Package*, is mainly designed to assist researchers and policy analysts interested in conducting distributive analysis with Stata. In particular, *DASP* is built to:

- Estimate the most popular statistics (indices, curves) used for the analysis of poverty, inequality, social welfare, and equity.
- Estimate the differences in such statistics.
- Estimate standard errors and confidence intervals by taking full account of survey design.
- Support distributive analysis on more than one database.
- Perform the most popular poverty and decomposition procedures.
- Check for the ethical robustness of distributive comparisons.
- Unify syntax and parameter use across various estimation procedures for distributive analysis.

For each *DASP* module, three types of files are provided:

- \*.ado: This file contains the program of the module
- \*.hlp: This file contains help material for the given module
- \*.dlg : This file allows the user to perform the estimation using the module's dialog box

The \*.dlg files make the *DASP* package very user-friendly and easy to learn. When these dialog boxes are used, the associated program syntax is also generated and showed in the review window. The user can save the contents of this window in a \*.do file to be subsequently used in another session.

## 2 DASP and Stata versions

DASP 3.0X requires

- Stata version 15.0 or higher

## 3 Installing and updating the DASP package

In general, the \*.ado files are saved in the following main directories:

Priority	Directory	Sources
1	UPDATES:	Official updates of Stata *.ado files
2	BASE:	*.ado files that come with the installed Stata software
3	SITE:	*.ado files downloaded from the net
4	PLUS:	..
5	PERSONAL:	Personal *.ado files

### 3.1 Installing DASP

Local installation	Internet installation
a. Download, then unzip the file <b>dasp302.zip</b> in the directory <b>c:/temp http://dasp.ecn.ulaval.ca/dasp302.zip</b> b. Make sure that you have <b>c:/dasp302/dasp.pkg</b> or <b>c:/temp/dasp302/stata.toc</b> (the directory <b>ca</b> be updated). c. In the Stata command windows, type the syntax: net from c:/temp/dasp302 net install dasp_p1.pkg, force replace net install dasp_p2.pkg, force replace net install dasp_p3.pkg, force replace net install dasp_p4.pkg, force replace net install dasp_p5.pkg, force replace net install dasp_p6.pkg, force replace addmenu profile.do _daspmenu	In the Stata command windows, type the syntax: net from http://dasp.ecn.ulaval.ca/dasp3 net install dasp_p1.pkg, force replace net install dasp_p2.pkg, force replace net install dasp_p3.pkg, force replace net install dasp_p4.pkg, force replace net install dasp_p5.pkg, force replace net install dasp_p6.pkg, force replace addmenu profile.do _daspmenu

Figure 1: Ouput of *net describe dasp*

```
. net from http://dasp.ecn.ulaval.ca/dasp302
=====
http://dasp.ecn.ulaval.ca/dasp302/
=====

Version      : Version   3.02
Date        : November 2021
Stata Version : Required 15.0 and higher

Author:
DASP is conceived by:

Dr. Abdelkrim Araar : aabd@ecn.ulaval.ca
Dr. Jean-Yves Duclos : jyves@ecn.ulaval.ca

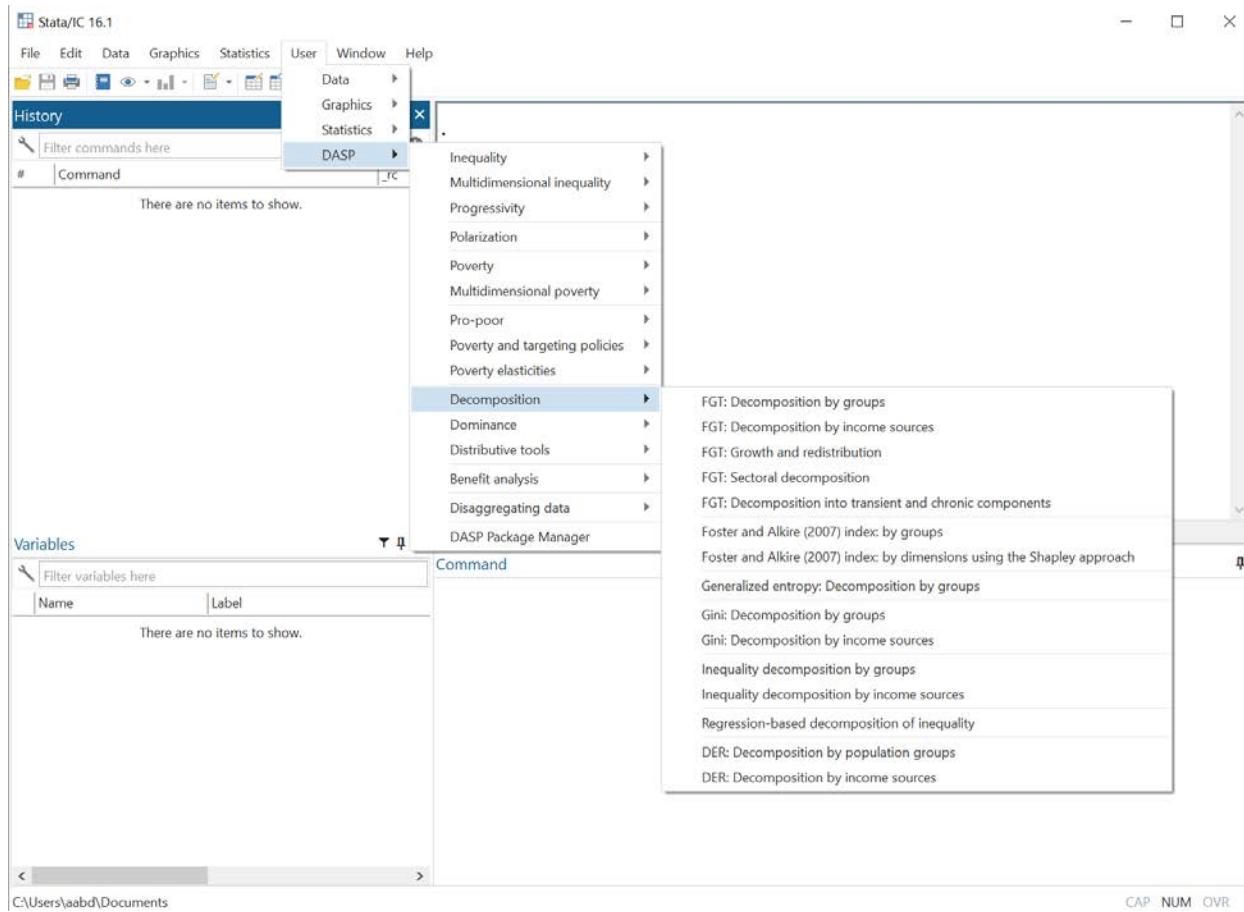
The next follwing sub-packages must be installed to run DASP.
=====

PACKAGES you could -net describe-:
dasp_p1          Distributive Analysis Stata Package: PART I
dasp_p2          Distributive Analysis Stata Package: PART II
dasp_p3          Distributive Analysis Stata Package: PART III
dasp_p4          Distributive Analysis Stata Package: PART IV
dasp_p5          Distributive Analysis Stata Package: PART V
dasp_p6          Distributive Analysis Stata Package: PART VI
=====
```

### 3.2 Adding the *DASP* submenu to Stata's main menu

With Stata 15 or higher, sub menus can be added to the menu item **User**.

Figure 2: DASP submenu



## 4 DASP and data files

DASP makes it possible to use simultaneously more than one data file. The user should, however, "initialise" each data file before using it with DASP. This initialisation is done by:

1. Labeling variables and values for categorical variables.
2. Initializing the sampling design with the command `svyset`.
3. Saving the initialized data file.

Users are recommended to consult appendices A, B and C.

## 5 Main variables for distributive analysis

VARIABLE OF INTEREST/WELFARE VARIABLE. This is the variable that usually captures living standards. It can represent, for instance, income per capita, expenditures per adult equivalent, calorie intake, normalised height-for-age scores for children, or household wealth.

SIZE VARIABLE. This refers to the "ethical" or physical size of the observation. For the computation of many statistics, we will indeed wish to consider how many relevant individuals (or statistical units) are found in a given observation.

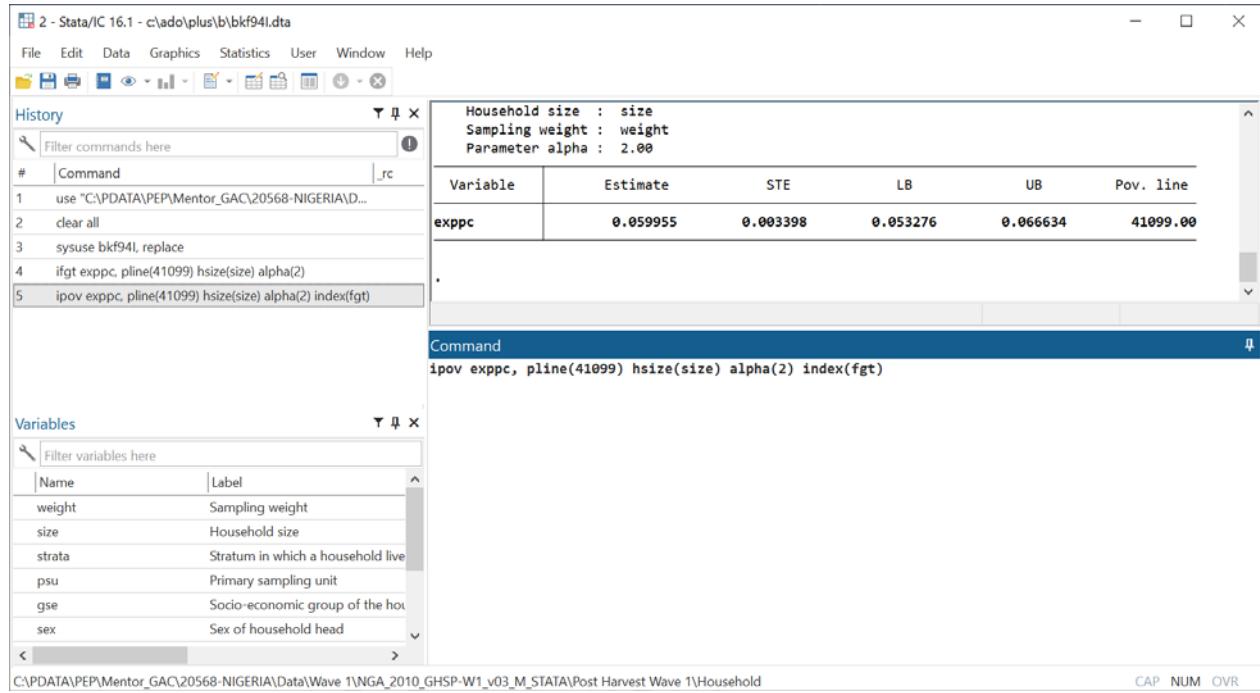
GROUP VARIABLE. (This should be used in combination with GROUP NUMBER.) It is often useful to focus one's analysis on some population subgroup. We might, for example, wish to estimate poverty within a country's rural area or within female-headed families. One way to do this is to force *DASP* to focus on a population subgroup defined as those for whom some GROUP VARIABLE (say, area of residence) equals a given GROUP NUMBER (say 2, for rural area).

SAMPLING WEIGHT. Sampling weights are the inverse of the sampling probability. This variable should be set upon the initialisation of the dataset.

## 6 How can DASP commands be invoked?

Stata commands can be entered directly into a command window:

Figure 3: Using DASP with a command window



An alternative is to use dialog boxes. For this, the command *db* should be typed and followed by the name of the relevant DASP module.

Example:

*db ipov*

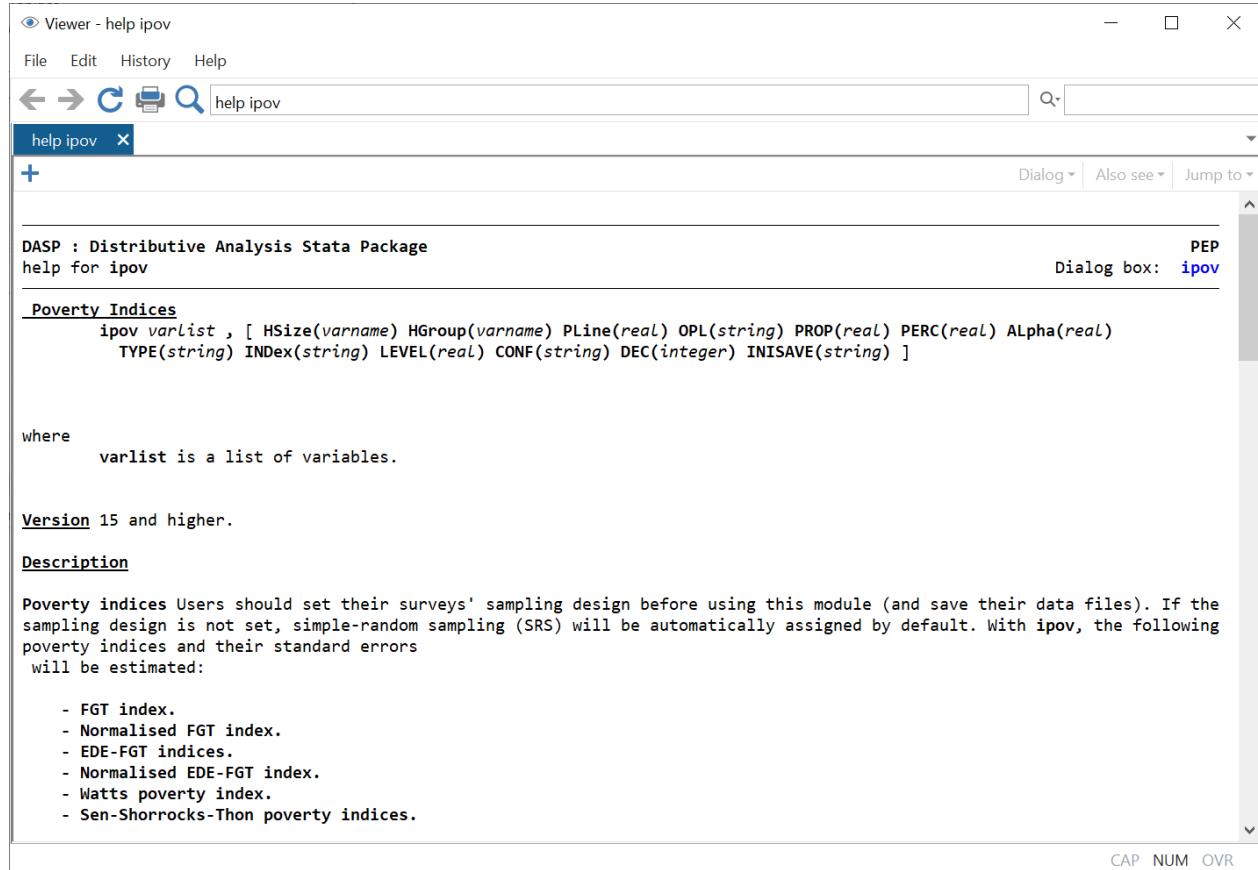
## 7 How can help be accessed for a given DASP module?

Type the command *help* followed by the name of the relevant *DASP* module.

Example:

*help ipov*

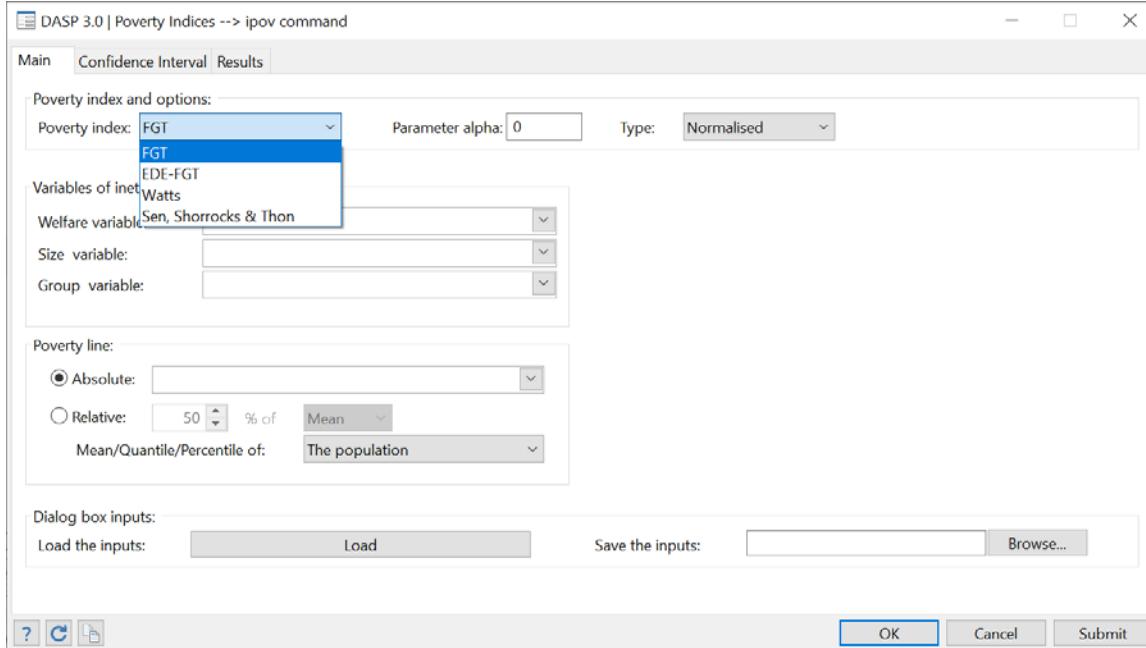
Figure 4: Accessing help on *DASP*



## 8 Applications and files in DASP

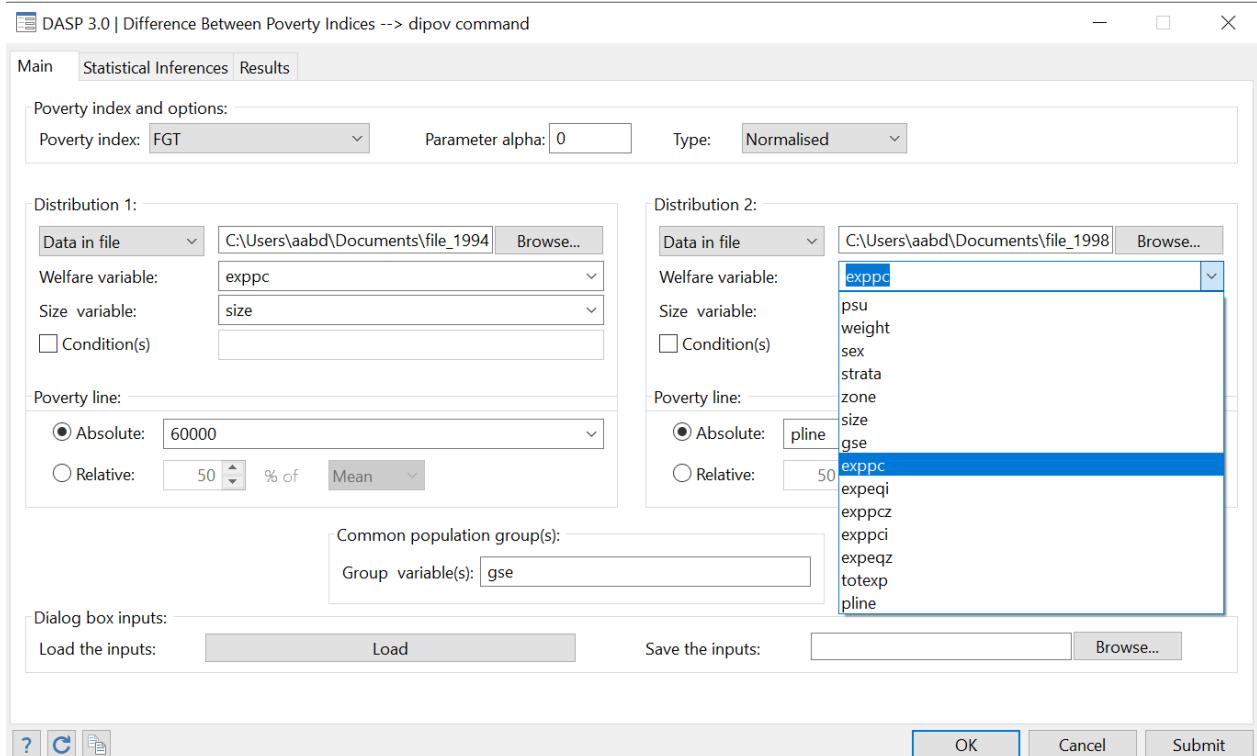
Two main types of applications are provided in *DASP*. For the first one, the estimation procedures require only one data file. In such cases, the data file in memory is the one that is used (or “loaded”); it is from that file that the relevant variables must be specified by the user to perform the required estimation.

Figure 5: Estimating FGT poverty with one distribution



For the second type of applications, two distributions are needed. For each of these two distributions, the user can specify the currently loaded data file (the one in memory) or one saved on disk.

**Figure 6: Estimating FGT poverty with two distributions**



Notes:

1. DASP considers two distributions to be statistically dependent (for statistical inference purposes) if the same data set is used (the same loaded data or data with the same path and filename) for the two distributions.
2. If the option DATA IN FILE is chosen, the keyboard must be used to type the name of the required variables.
3. Lists of varnames of DATA IN FILE are automatically loaded.

## 9 Basic Notation

The following table presents the basic notation used in *DASP*'s user manual.

Symbol	Indication
$y$	variable of interest
$i$	observation number
$y_i$	value of the variable of interest for observation i
$hw$	sampling weight
$hw_i$	sampling weight for observation i
$hs$	size variable
$hs_i$	size of observation i (for example the size of household i)
$w_i$	$hw_i * hs_i$
$hg$	group variable
$hg_i$	group of observation i.
$w_i^k$	$sw_i^k = sw_i$ if $hg_i = k$ , and 0 otherwise.
$n$	sample size

For example, the mean of  $y$  is estimated by *DASP* as  $\hat{\mu}$ :

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

## 10 DASP and poverty indices

### 10.1 Poverty indices (ipov)

The **ipov** module can be used to estimate different poverty measurements we introduce below.

#### FGT index

The non-normalised Foster-Greer-Thorbecke or FGT index is estimated as

$$\hat{P}(z; \alpha) = \frac{\sum_{i=1}^n w_i (z - y_i)_+^\alpha}{\sum_{i=1}^n w_i}$$

where  $z$  is the poverty line and  $x_+ = \max(x, 0)$ .

The usual normalised FGT index is estimated as

$$\hat{\bar{P}}(z; \alpha) = \hat{P}(z; \alpha)/(z)^\alpha$$

### **EDE-FGT index**

The “equally-distributed-equivalent” (EDE)-FGT EDE-FGT index is estimated as

$$\widehat{EDE}(P(z; \alpha)) = (\hat{P}(z; \alpha))^{1/\alpha} \quad \text{for } \alpha > 0$$

### **Watts poverty index**

The Watts poverty index is estimated as

$$\hat{P}(z) = \frac{\sum_{i=1}^q w_i (\ln(z/y_i))}{\sum_{i=1}^n w_i}$$

where  $z$  is the poverty line and  $q$  is the number of poor people.

### **Sen-Shorrocks-Thon poverty index**

The Sen-Shorrocks-Thon poverty index is estimated as

$$\hat{P}(z) = HP^*(z, \alpha)[1 + G_g^*]$$

where  $z$  is the poverty line,  $H$  is the headcount,  $P^*(z, \alpha)$  is the poverty gap estimated at the level of the poor group and  $G_g^*$  is the Gini index of poverty gaps  $(z - y)_+/z$ .

Note that the following points apply to the different indices:

- There exists three ways of fixing the poverty line:
  - 1- Setting a deterministic poverty line;
  - 2- Setting the poverty line to a proportion of the mean; or
  - 3- Setting the poverty line to a proportion of a quantile  $Q(p)$ .
- The user can choose the value of parameter  $\alpha$  (for the FGT and EDE-FGT indices).
- The user can select more than one variable of interest at a time. For example, poverty can be estimated by simultaneously using *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorial group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.

Interested users are encouraged to consider the exercises that appear in Section 23.4.

## 10.2 Differences between poverty indices (dipov)

This module estimates the differences between the poverty indices of two distributions.

For each distribution:

- There exists three ways of fixing the poverty line:
  - 1- Setting a deterministic poverty line;
  - 2- Setting the poverty line to a proportion of the mean; or
  - 3- Setting the poverty line to a proportion of a quantile  $Q(p)$ .
- One variable of interest should be selected at a time.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.
- The user can choose a separate value for parameter  $\alpha$ .

Interested users are encouraged to consider the exercises that appear in Section 23.5.

- The user can select more than one variable of interest at a time. For example, poverty can be estimated by simultaneously using *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorial group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.

## 10.3 DASP and multidimensional poverty indices (imd pov)

The general form of an additive multidimensional poverty index is:

$$P(X, Z) = \frac{\sum_{i=1}^n w_i p(X_i, Z)}{\sum_{i=1}^n w_i}$$

where  $p(X_i, Z)$  is individual  $i$ 's poverty function (with vector of attributes  $X_i = (x_{i,1}, \dots, x_{i,J})$  and vector of poverty lines  $Z = (z_1, \dots, z_J)$ ), determining  $i$ 's contribution to total poverty  $P(X, Z)$ .

[1] Chakravarty et al. (1998) index

$$p(X_i, Z) = \sum_{j=1}^J a_j \left( \frac{z_j - x_{i,j}}{z_j} \right)_+^\alpha$$

[2] Extended Watts index

$$p(X_i, Z) = \sum_{j=1}^J a_j \ln \left( \frac{z_j}{\min(z_j; x_{i,j})} \right)$$

[3] Multiplicative extended FGT index

$$p(X_i, Z) = \prod_{j=1}^J \left( \frac{z_j - x_{i,j}}{z_j} \right)_+^{\alpha_j}$$

[4] Tsui (2002) index

$$p(X_i, Z) = \prod_{j=1}^J \left( \frac{z_j}{\min(z_j; x_{i,j})} \right)^{b_j} - 1$$

[5] Intersection headcount index

$$p(X_i, Z) = \prod_{j=1}^J I(z_j > x_{i,j})$$

[6] Union headcount index

$$p(X_i, Z) = 1 - \prod_{j=1}^J I(z_j < x_{i,j})$$

[7] Bourguignon and Chakravarty (2003) index

$$\frac{1}{N} \sum_{i=1}^N \left( \sum_{j=1}^J \left[ \beta_j^{\gamma/\alpha} C_{j,i} \right]^{\alpha/\gamma} \right)^p$$

where:

$$C_{j,i} = \left( \frac{z_1 - x_{j,i}}{z_j} \right)_+^\gamma$$

[8] Alkire and Foster (2011) index

$$p(\alpha, X_i, Z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{J} \sum_{j=1}^J w_j \left( \frac{z_j - x_{i,j}}{z_j} \right)_+^\alpha I(d_i \geq d_c)$$

where  $\sum_{j=1}^J w_j = J$  and  $d_i$  denotes the number of dimensions in which the individual  $i$  is deprived.  $d_c$  denotes the normative dimensional cut-off.

The modules presented above can be used to estimate the multidimensional poverty indices as well as their standard errors.

- The user can select among the seven multidimensional poverty indices.
- The number of dimensions can be selected (1 to 10). With the option `pinof(vars)`, the user indicates more than 10 dimensions (see the help of the command `imdpov`).
- If applicable, the user can choose parameter values relevant to a chosen index.
- A group variable can be used to estimate the selected index at the level of a categorical group.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided, and the level of confidence used can be changed.
- The results are displayed with 3 decimals; this can also be changed.

Users are encouraged to consider the exercises that appear in Section 23.6

## 10.4 Multiple overlapping deprivation analysis (MODA) indices

The **imoda** DASP module produces a series of multidimensional poverty indices to show the incidence of deprivation in each dimension. Further, this application estimates the incidence of multi-deprivation in the different combinations of dimensions. In this application, the number of dimensions is set to three. Further, the multidimensional poverty is measured by the headcount (union and intersection headcount indices) and the Alkire and Foster (2011) Mo index for different levels of the dimensional cut-off.

- The number of dimensions is three.
- A group variable can be used to estimate the MODA indices at the level of a categorical group.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided, and the level of confidence used can be changed.
- The results are displayed with 3 decimals; this can be also changed.

## Illustrative application

To illustrate the MODA DASP module, we start by constructing three dimensions of well-being using the Cameroonian Household Survey (ECAM II: Enquête Camerounaise Auprès des Ménages), which is conducted by the National Institute of Statistics in 2001. This is a national survey with a sample of about 11,000 households selected randomly using two stages in the urban areas and three for the rural areas. Besides the detailed information on household expenditures, this sample contains rich information about the non-monetary dimensions of well-being, such as the access to public goods. For our application, we focus on three dimensions of well-being, which are the housing, education, and health basic infrastructures.

As is well-known, there is no unique indicator as a measure for a given dimension of well-being at the individual level. In general, one has to quantify this level starting from a set of primary indicators, which will be strongly related to the dimension that they represent. Starting from the fact that all used indicators in this example are categorical, we propose to use the Multiple Correspondence Analysis (MCA) technique to estimate the individual normalised scores, and this, for each of the three retained dimensions of well-being. One can recall here that the MCA is the application of the simple correspondence analysis to multivariate categorical data, coded as an indicator matrix or a Burt matrix. The level of well-being of the individual  $i$  for a given dimension is quantified as follows:

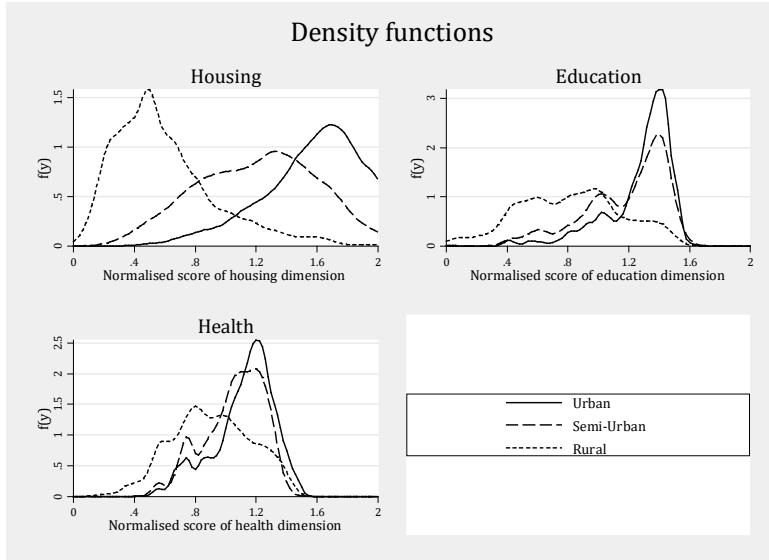
$$W_i = \frac{\sum_{k=1}^K \sum_{j_k=1}^{J_k} w_{j_k} I_{i,j_k}}{K}$$

where  $K$  is the number of categorical variables,  $J_k$  the number of categories for indicator  $k$ ,  $I_{i,j_k}$  the binary indicator taking 1 if the individual  $i$  having the category  $j_k$  and  $w_{j_k}$  is the normalised first axis score of the category  $j_k$ <sup>1</sup>. In resume, after reducing the information from the  $\sum_k J_k$  dimensions to the

---

<sup>1</sup>Precisely, the normalised weight refers to the score on the first axis, normalised by the square root of its correspondent

fewest one, the reduced space preserves the main disparities in well-being. Precisely, the main part of this disparity is projected on the first axis -factor- of the reduced space. This is the reason for which one can use the categorical scores of the first axis as categorical weights. In figure 1, we plot the density curves of each of the three dimensions of well-being according to the area where the household lives. As expected, urban households have better scores in housing, education and health infrastructures compared to those that live in semi-urban and rural areas. However, how is extent of correlation and overlapping between the dimensional deprivations?



```
. corr housing education health [aw= weight ]
(sum of wgt is 3.1209e+06)
(obs=10991)
```

	housing	educat~n	health
housing	1.0000		
education	0.6627	1.0000	
health	0.3995	0.3954	1.0000

To answer this question, we propose to produce the MODA results using the DASP package.

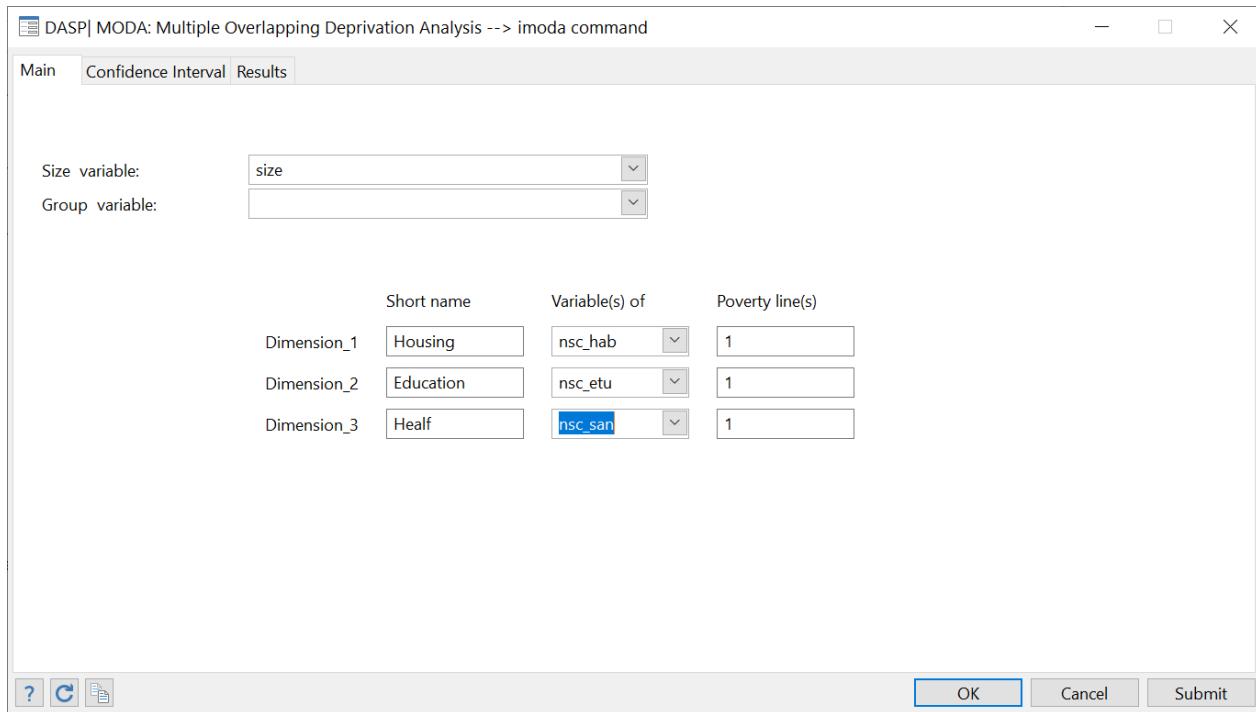
To invoke the dialogue box of the moda application, we type the command line.

*db imoda*

---

eigenvalue.

**Figure 7: The IMODA dialog box**



- Of course, the application enables the user to select any three dimensions of well-being.

```
. imoda nsc_hab nsc_etu nsc_san, hsize(size) pl1(1) pl2(1) pl3(1) sn1(Housing) sn2(Education) sn3(Healf)
Multiple Overlapping Deprivation Analysis
Household size      : size
Sampling weight     : weight
Notes:
- Dim1: Housing
- Dim2: Education
- Dim3: Healf
```

Dimensions	Headcount (H)		M0		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	55.42	55.42	55.42	---	---
Dim2	47.58	47.58	47.58	---	---
Dim3	46.05	46.05	46.05	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	62.58	40.42	51.50	40.42	---
Dim1 - Dim3	67.39	34.07	50.73	34.07	---
Dim2 - Dim3	64.12	29.51	46.81	29.51	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	71.41	26.37	49.68	42.97	26.37

- Showing more detailed results (use the option detail(1)).

```
. imoda housing education health, hsize(size) pl1(1) pl2(1) pl3(1) sn1(Housing) sn2(Education) sn3(Health) detail(1)
Multiple Overlapping Deprivation Analysis
Household size      : size
Sampling weight     : weight
Notes:
- Dim1: Housing
- Dim2: Education
- Dim3: Health
```

Dimensions	Headcount (H)		M0		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	55.42	55.42	55.42	---	---
Dim2	47.58	47.58	47.58	---	---
Dim3	46.05	46.05	46.05	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	62.58	40.42	51.50	40.42	---
Dim1 - Dim3	67.39	34.07	50.73	34.07	---
Dim2 - Dim3	64.12	29.51	46.81	29.51	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	71.41	26.37	49.68	42.97	26.37

Detailed results for the headcount index (H) and two dimensions:

Dimensions	Union: D1 OR D2	Intersection: D1 and D2	D1 and not D2	D2 and not D1
D1:Dim1 // D2:Dim2	62.58	40.42	14.99	7.16
D1:Dim1 // D2:Dim3	67.39	34.07	21.34	11.97
D1:Dim2 // D2:Dim3	64.12	29.51	18.07	16.54

Showing standard errors by taking into account the full information about the sampling design (adding the option dste(1)).

```
. imoda nsc_hab nsc_etu nsc_san, hsize(size) pl1(1) pl2(1) pl3(1) sn1(Housing) sn2(Education) sn3(Healf) detail(1)
Multiple Overlapping Deprivation Analysis
Household size      : size
Sampling weight     : weight
Notes:
- Dim1: Housing
- Dim2: Education
- Dim3: Healf
```

Dimensions	Headcount (H)		M0		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	55.42	55.42	55.42	---	---
Dim2	47.58	47.58	47.58	---	---
Dim3	46.05	46.05	46.05	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	62.58	40.42	51.50	40.42	---
Dim1 - Dim3	67.39	34.07	50.73	34.07	---
Dim2 - Dim3	64.12	29.51	46.81	29.51	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	71.41	26.37	49.68	42.97	26.37

Detailed results for the headcount index (H) and two dimensions:

Dimensions	Union: D1 OR D2	Intersection: D1 and D2	D1 and not D2	D2 and not D1
D1:Dim1 // D2:Dim2	62.58	40.42	14.99	7.16
D1:Dim1 // D2:Dim3	67.39	34.07	21.34	11.97
D1:Dim2 // D2:Dim3	64.12	29.51	18.07	16.54

- Producing results by population groups (option: hgroup(area))

```
. imoda nsc_hab nsc_etu nsc_san, hsiz(size) hgroup(area) p11(1) p12(1) p13(1) sn1(Housing) sn2(Education) sn3(Health) detail(1) dste(1)
Multiple Overlapping Deprivation Analysis
Household size : size
Sampling weight : weight
Group variable : area
Notes:
- Dim1: Housing
- Dim2: Education
- Dim3: Health
Results for the whole population
```

Dimensions	Headcount (H)		M8		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	55.42	55.42	55.42	---	---
	0.74	0.74	0.74	---	---
Dim2	47.58	47.58	47.58	---	---
	0.77	0.77	0.77	---	---
Dim3	46.05	46.05	46.05	---	---
	0.77	0.77	0.77	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	62.58	48.42	51.58	48.42	---
	0.69	0.79	0.66	0.79	---
Dim1 - Dim3	67.39	34.87	59.73	34.87	---
	0.65	0.77	0.61	0.77	---
Dim2 - Dim3	64.12	29.51	46.81	29.51	---
	0.68	0.72	0.59	0.72	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	71.41	26.37	49.68	42.97	26.37
	0.61	0.71	0.56	0.65	0.71

Detailed results for the headcount index (H) and two dimensions:

Dimensions	Union: D1 OR D2	Intersection: D1 and D2	D1 and not D2	D2 and not D1
D1:Dim1 // D2:Dim2	62.58 0.69	48.42 0.79	14.99 0.68	7.16 0.36
D1:Dim1 // D2:Dim3	67.39 0.65	34.87 0.77	21.34 0.69	11.97 0.45
D1:Dim2 // D2:Dim3	64.12 0.68	29.51 0.72	18.87 0.67	16.54 0.61

- Results for the group: Urban

Notes:  
- Dim1: Housing  
- Dim2: Education  
- Dim3: Health

Dimensions	Headcount (H)		M8		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	11.23 0.49	11.23 0.49	11.23 0.49	---	---
Dim2	15.83 0.59	15.83 0.59	15.83 0.59	---	---
Dim3	24.98 0.72	24.98 0.72	24.98 0.72	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	21.79 0.66	5.27 0.35	13.53 0.44	5.27 0.35	---
Dim1 - Dim3	31.72 0.77	4.48 0.32	18.10 0.46	4.48 0.32	---
Dim2 - Dim3	34.18 0.79	6.70 0.48	28.40 0.50	6.70 0.40	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	38.04 0.81	2.46 0.24	17.34 0.41	8.51 0.38	2.46 0.24

Detailed results for the headcount index (H) and two dimensions:

Dimensions	Union: D1 OR D2	Intersection: D1 and D2	D1 and not D2	D2 and not D1
D1:Dim1 // D2:Dim2	21.79 0.66	5.27 0.35	5.96 0.36	10.56 0.50
D1:Dim1 // D2:Dim3	31.72 0.77	4.48 0.32	6.75 0.39	20.49 0.67
D1:Dim2 // D2:Dim3	34.18 0.79	6.70 0.40	9.13 0.46	18.28 0.64

- Results for the group: Rural

Notes:  
- Dim1: Housing  
- Dim2: Education  
- Dim3: Health

Dimensions	Headcount (H)		M8		
	Union	Intersection	Cut-off = 1	Cut-off = 2	Cut-off = 3
<b>- One dimension:</b>					
Dim1	88.69 0.77	88.69 0.77	88.69 0.77	---	---
Dim2	71.49 1.09	71.49 1.09	71.49 1.09	---	---
Dim3	61.91 1.16	61.91 1.16	61.91 1.16	---	---
<b>- Two dimensions:</b>					
Dim1 - Dim2	93.29 0.61	66.89 1.13	80.89 0.74	66.89 1.13	---
Dim1 - Dim3	94.25 0.54	56.35 1.18	75.30 0.71	56.35 1.18	---
Dim2 - Dim3	86.72 0.73	46.68 1.17	66.70 0.77	46.68 1.17	---
<b>- Three dimensions:</b>					
Dim1 - Dim2 - Dim3	96.54 0.49	44.37 1.16	74.03 0.62	68.92 0.82	44.37 1.16

Detailed results for the headcount index (H) and two dimensions:

Dimensions	Union: D1 OR D2	Intersection: D1 and D2	D1 and not D2	D2 and not D1
D1:Dim1 // D2:Dim2	93.29 0.61	66.89 1.13	21.88 1.00	4.60 0.51
D1:Dim1 // D2:Dim3	94.25 0.54	56.35 1.18	32.34 1.12	5.56 0.59
D1:Dim2 // D2:Dim3	86.72 0.73	46.68 1.17	24.41 1.08	11.23 0.95

- Deprivations in housing and education are more correlated in rural area, while deprivations in education and health are more correlated in urban area.
- Other conclusions can be found by analysing the results.

## 11 DASP, poverty and targeting policies

### 11.1 Poverty and targeting by population groups

The per capita dollar impact of a marginal addition of a constant amount of income to everyone within a group  $k$  – called Lump-Sum Targeting (LST) – on the FGT poverty index  $P(k, z; \alpha)$ , is as follows:

$$LST = \begin{cases} -\alpha P(k, z; \alpha - 1) & \text{if } \alpha \geq 1 \\ -f(k, z) & \text{if } \alpha = 0 \end{cases}$$

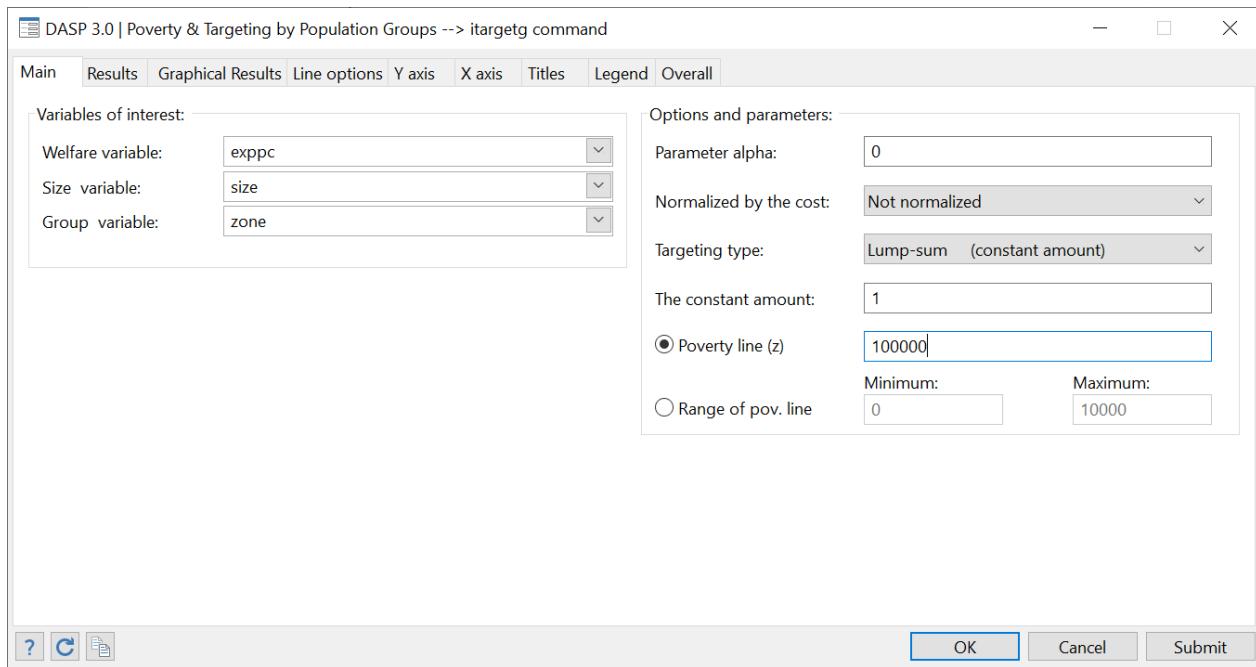
where  $z$  is the poverty line,  $k$  is the population subgroup for which we wish to assess the impact of the income change, and  $f(k, z)$  is the density function of the group  $k$  at level of income  $z$ . The per capita dollar impact of a proportional marginal variation of income within a group  $k$ , called Inequality Neutral Targeting, on the FGT poverty index  $P(k, z; \alpha)$  is as follows:

$$INT = \begin{cases} \alpha \frac{P(k, z; \alpha) - zP(k, z; \alpha - 1)}{\mu(k)} & \text{if } \alpha \geq 1 \\ -\frac{zf(k, z)}{\mu(k)} & \text{if } \alpha = 0 \end{cases}$$

The module **itargetg** allows to:

- Estimate the impact of marginal change in income of the group on poverty of the group and that of the population.
- Select the design of change, constant or proportional to income to keep inequality unchanged.
- Draw curves of impact according for a range of poverty lines.
- Draw the confidence interval of impact curves or the lower or upper bound of confidence interval.
- Etc.

**Figure 8: Poverty and the targeting by population groups**



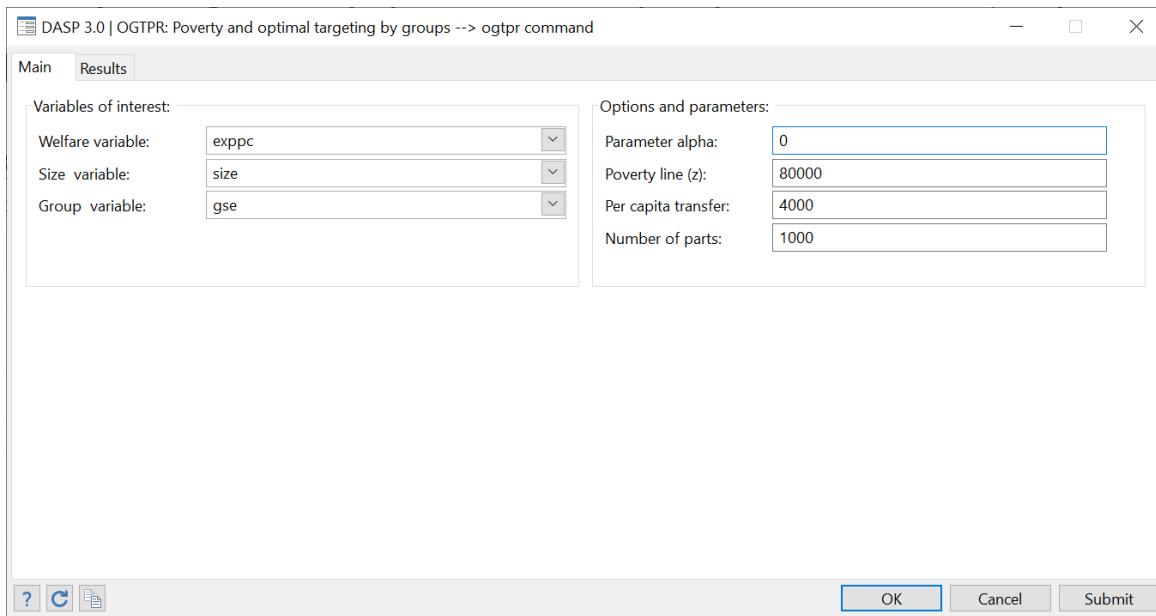
#### Reference

DUCLOS, J.-Y. AND A. ARAAR (2006): Poverty and Equity Measurement, Policy, and Estimation with DAD, Berlin and Ottawa: Springer and IDRC. (sec. 12.1)

## 11.2 Poverty and targeting by population group with a fixed budget (*ogtpr*)

The *ogtpr* module is designed to provide a numerical solution to the problem of reduction in FGT poverty when the total transfer budget is predetermined and the transfer amount is universal at the group level (i.e., all members of a population group receive the same amount). The numerical algorithm used was developed by Araar and Tiberti (2020).

**Figure 9: Poverty and targeting by population group in the case of a fixed budget**



To better understand the module, let us discuss the example presented in Figure 8. First, we assume that the welfare variable is per capita expenditures. The social planner:

1. Must find the optimal combination of group transfers to reduce the headcount at the population level as much as possible;
2. Must give every person in the population 4000 (monetary units) because the government's total transfer budget is equivalent to a transfer of 4000 (monetary units) per capita;
3. Is free to fix the group transfer amount when population groups are based on the socio-economic group of the household head (variable **gse**: seven population groups).

Fortunately, the algorithm can do this difficult job for the planner, and the results are as follows:

```
. ogtpc exppc, hgroup(gse) hsize(size) alpha(0) pline(80000) trans(4000) part(1000) ered(1)
Sequence ...1: Remaining p.c. budget 3952.326 over 4000.000
Sequence ...2: Remaining p.c. budget 3878.941 over 4000.000
Sequence ...3: Remaining p.c. budget 3807.127 over 4000.000
Sequence ...4: Remaining p.c. budget 0.000 over 4000.000
```

```
Optimal Targeting of Groups for Poverty Reduction
Number of observations : 8478
Time of computation : 5.73 second(s)
Per capita transfer : 4000.00
Used per capita transfer : 4000.00
Household size : size
Sampling weight : weight
Group variable : gse
Parameter alpha : 0.00
```

Group	Fgt Index	Population Share	Optimal G.P.C. Transfer	Optimal P.C. Transfer
Wage-earner (public sector)	0.075	4.140	0.000	0.000
Wage-earner (private sector)	0.139	2.904	0.000	0.000
Artisan or trader	0.155	5.580	0.000	0.000
Other type of earner	0.355	0.569	0.000	0.000
Crop farmer	0.492	16.781	22971.742	3854.801
Subsistence farmer	0.606	65.355	0.000	0.000
Inactive	0.463	4.672	3107.906	145.199

#### Total Poverty Reduction

Variable	Estimate	Std. Err.	t	P> t	[95% Conf. interval]	Pov. line
exppc	.5180611	.0109517	47.3042	0.0000	.4965334 .5395888	80000
exppc_tr	.4765408	.0112242	42.4565	0.0000	.4544774 .4986042	80000
diff.	-.0415202	.0046131	-9.0005	0.0000	-.0505882 -.0324522	---

```
- Reduction with imperfect targeting (in %) : -4.152
- Reduction with perfect targeting (in %) : -28.247
- The quality of the targeting indicator (in %): 14.699
```

#### Targeting by transfers and poverty status

Poor	Targeted		Total
	No	Yes	
No	37.17	11.03	48.19
Yes	41.38	10.43	51.81
Total	78.55	21.45	100.00

The first table shows:

- The initial poverty level headcount of each group;
- The population share of each group (as a percentage);
- The optimal transfer amount for each group (the amount that each member within the group should receive);
- The population per transfer (the cost of the group transfer at the population level); of course, the total of the rows in this column is equal to the per-capita budget (4000 monetary units).

The algorithm suggests to give transfers only to groups 5 and 7 (22971.7 for each person in Group 5 and 145.2 for each person in Group 7).

After these optimal group transfers are determined with the **gse** variable, the headcount decreases from 51.81% to 47.65%, which represents a 4.152% reduction in poverty (results of the second table).

Let's assume that the planner can give each person a distinct amount. This case is what we call perfect targeting. With perfect targeting and the same budget, the total reduction in poverty can reach 28.247% instead of 4.152%. Thus, the **gse** tool improves the quality of targeting by  $(4.152/28.247)*100 = 14.699\%$ .

#### Reference

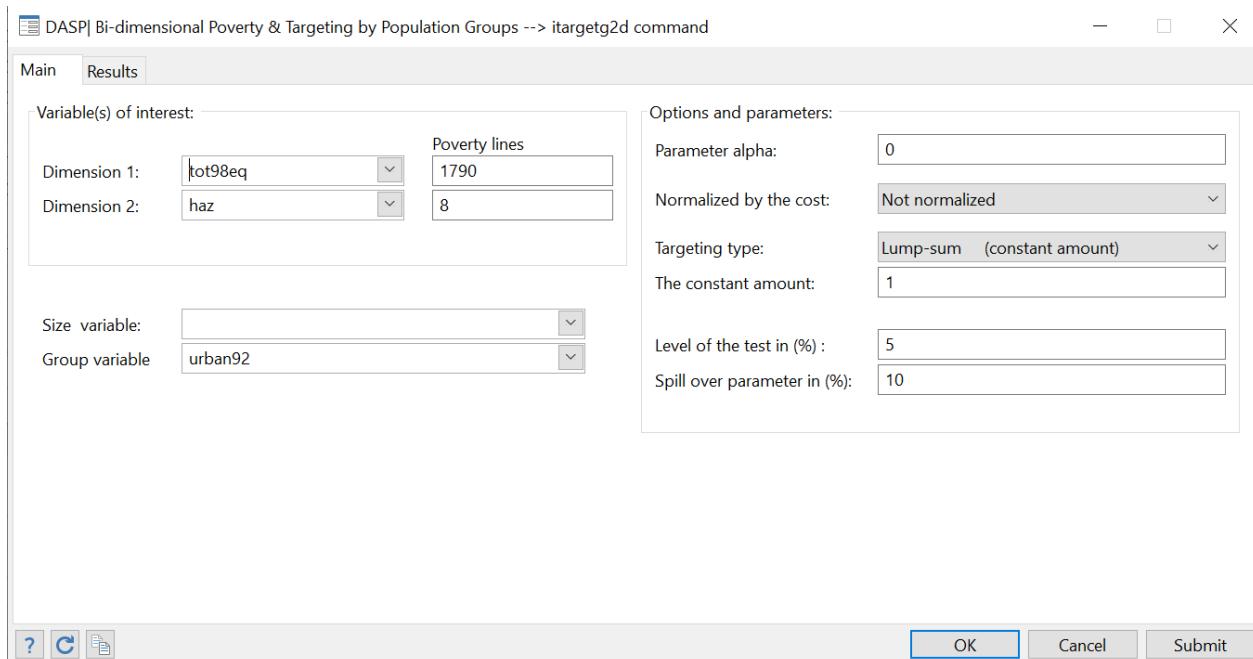
Araar Abdelkrim and Luca Tiberti (2020), Optimal Population Group Targeting and Poverty Reduction, mimeo. [http://dasp.ecn.ulaval.ca/dasp3/refs/Optimal\\_Targeting\\_and\\_Poverty\\_Reduction\\_april2020.pdf](http://dasp.ecn.ulaval.ca/dasp3/refs/Optimal_Targeting_and_Poverty_Reduction_april2020.pdf)

## 11.3 Bi-dimensional poverty and targeting by population group (*itargetg2d*)

The *itargetg2d* module is designed to estimate optimal group targeting in order to reduce bi-dimensional poverty by considering the joint distribution of deprivations. Duclos, Tiberti and Araar (2018) used Duclos, Sahn and Younger's (2006) multidimensional poverty index to formally derive efficient targeting rules in a normative multidimensional poverty setting. In accordance with Duclos, Tiberti and Araar (2018), the module considers three channels through which targeting may affect multidimensional poverty: (1) a direct effect on the targeted dimension, (2) an indirect effect on the level of joint deprivation, and (3) a spillover effect on the other dimensions. The user can estimate the impact of a lump-sum or a proportional transfer on poverty. They can also indicate the spillover parameter and normalise the estimated poverty reduction by the cost of the transfer to be able to compare effects across different population groups.

As shown in Duclos, Tiberti and Araar (2018), the total effect can be decomposed into different components. The *itargetg2d* module makes it possible to estimate the different components. It also provides an ordinal ranking of the groups based on the statistical significance of the differences between the group impacts. The *itargetg2d* module has a dialog box as well as a Stata help file.

**Figure 10: Bi-dimensional poverty and targeting by population group**



This is an example; *it is strongly recommended to consult Duclos et al. (2018) to familiarise yourself with the different results.*

```

. itargetg2d tot98eq haz, alpha(0) pline1(1790) pline2(8) hgroup(urban92) constam(1) slevel(5) so(10)

Targeting population groups and poverty
Targeting groups : Groups => urban92
Targeting scheme : Lump-sum (constant)
Normalized by cost : no
Parameter alpha : 0.00
Parameter Spill-Over: 0.10
Poverty line 1 : 1790.00
Poverty line 2 : 8.00

```

Group	Population Share	Multiplicative Index	Impact on Population
Rural	0.847131431	0.517788231	-0.006285515
	0.009406213	0.015651515	0.000625260
Urban	0.152868554	0.223277912	-0.002963044
	0.009406213	0.036942225	0.000328483
Population	1.000000000	0.472766876	-0.009244086
	0.000000000	0.014501242	0.000705459

Impact based on the component  $-\alpha/z^* [P_1(\alpha-1)]$

Rank	Code	Label	Estimated Impact	Standard Error	Test of difference Signif. = 5%
1	0	Rural	-0.000320951	0.000024059	0<-:1
2	1	Urban	-0.000054396	0.000006905	1<-

Impact based on the component  $-\alpha/z^* [P_1(\alpha-1)*P_2(\alpha)]$

Rank	Code	Label	Estimated Impact	Standard Error	Test of difference Signif. = 5%
1	0	Rural	-0.000203879	0.000014937	0<-:1
2	1	Urban	-0.000024420	0.000003918	1<-

Total impact without spillover :  $-\alpha/z^* [P_1(\alpha-1)P_2(\alpha) + Cov(.)]$

Rank	Code	Label	Estimated Impact	Standard Error	Test of difference Signif. = 5%
1	0	Rural	-0.000144731	0.000015408	0<-:1
2	1	Urban	-0.000026168	0.000004341	1<-

Total impact with spillover

Rank	Code	Label	Estimated Impact	Standard Error	Test of difference Signif. = 5%
1	0	Rural	-0.006285515	0.000625260	0<-:1
2	1	Urban	-0.002963044	0.000328483	1<-

In the above example, we have a monetary dimension (approximated by per capita expenditures) and a health dimension (approximated by height-for-age z-scores). Furthermore, we use the Vietnam Living Standard Survey 1998 to estimate optimal targeting for a lump-sum transfer of \$1 when we consider two groups (urban and rural areas). The multiplicative poverty index is estimated for  $\alpha = 0$ . The first table provides the results and shows that per dollar spent, it is better to prioritise urban areas. In the second table, we focus only on the monetary dimension and find that although the unidimensional poverty reduction is greater when rural areas are targeted, the results are not statistically significant. When we also consider the health indicator (see the third table), rural areas should be statistically preferred. Unlike in the previous tables, once joint deprivation is considered, a larger – though not statistically significant –

impact is achieved when urban areas are targeted. The reversal in the results is explained by the relatively high density level around the two poverty lines in urban areas as compared to rural areas. In other words, in urban areas, health poverty is greater for those around the consumption poverty line; the correlation between the monetary and health dimensions is larger. The spill-over effect, which it is related to the impact of the monetary transfer on health, makes targeting urban areas statistically preferable to reduce our bi-dimensional poverty indicator in Vietnam more quickly. From this example, we learned how the priority ranking for targeting can be by a shift from a unidimensional setting to a multidimensional one and about the role each of the components included in the multidimensional indicator plays.

## References

- Duclos, J. Y., Tiberti, L., & Araar, A. (2018). Multidimensional Poverty Targeting. *Economic Development and Cultural Change*, 66(3), 519-554. (An earlier version was published as Cahier de recherche 1339, CIRPEE.)
- Duclos, J. Y., Sahn, D. E., & Younger, S. D. (2006). "Robust multidimensional poverty comparisons". *The Economic Journal*, 116(514), 943-968.

## 11.4 Poverty and targeting by income components

### **Proportional change per 100% of component**

Assume that total income  $Y$  is the sum of  $J$  income components, with  $Y = \sum_{j=1}^J \lambda_j y_j$  and where  $c$  is a factor that multiplies income component  $y_j$  and that can be subject to growth. The derivative of the normalised FGT index with respect to  $\lambda_j$  is given by

$$\left. \frac{\partial \bar{P}(z, \alpha)}{\partial \lambda_j} \right|_{\lambda_j=1, j=1 \dots J} = -CD_j(z, \alpha)$$

where  $CD_j$  is the Consumption dominance curve of component  $j$ .

### **Change per \$ of component**

The per capita dollar impact of growth in the  $j^{th}$  component on the normalised FGT index of the  $k^{th}$  group is as follows:

$$\frac{\frac{\partial P(z, \alpha)}{\partial y^j}}{\frac{\partial \mu_j}{\partial y^j}} = -\overline{CD}^j(z, \alpha)$$

where  $\overline{CD}^j$  is the normalised consumption dominance curve of the component  $j$ .

### **Constant change per component**

Simply we assume that the change concerns the group with component level greater than zero. Thus, this is similar to targeting by the nonexclusive population groups.

The module **itargetc** allows to:

- Estimate the impact of marginal change in income component on poverty;
- Select the option normalised or non-normalised by the average of component;
- Select the design of change, constant (lump sum) or proportional to income to keep inequality unchanged;
- Draw curves of impact according for a range of poverty lines;
- Draw the confidence interval of impact curves or the lower or upper bound of confidence interval;
- Etc.

### Reference

DUCLOS, J.-Y. AND A. ARAAR (2006): Poverty and Equity Measurement, Policy, and Estimation with DAD, Berlin and Ottawa: Springer and IDRC. (sec. 12)

## 12 Marginal poverty impacts and poverty elasticities

### 12.1 FGT elasticity's with respect to the average income growth (efgtgr).

The overall growth elasticity (GREL) of poverty, when growth comes exclusively from growth within a group  $k$  (namely, within that group, inequality neutral), is given by:

$$GREL = \begin{cases} -\frac{zf(k, z)}{F(z)} & \text{if } \alpha = 0 \\ \alpha \frac{\bar{P}(k, z; \alpha) - \bar{P}(k, z; \alpha - 1)}{\bar{P}(z, \alpha)} & \text{if } \alpha \geq 1 \end{cases}$$

where  $z$  is the poverty line,  $k$  is the population subgroup in which growth takes place,  $f(k, z)$  is the density function at level of income  $z$  of group  $k$ , and  $F(z)$  is the headcount.

Araar, Abdelkrim and Jean-Yves Duclos, (2007), [Poverty and inequality components: a micro framework](#), Working Paper: 07-35. CIRPEE, Department of Economics, Université Laval.

Kakwani, N. (1993) "Poverty and economic growth with application to Côte d'Ivoire", Review of Income and Wealth, 39(2): 121:139.

- To estimate the FGT elasticity's with respect average income growth the group or the whole population;
- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided, and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

## 12.2 FGT elasticities with respect to average income growth with different approaches (*efgtgro*).

The overall growth elasticity of poverty is estimated using one approach among the following list:

- The counterfactual approach;
- The marginal approach;
- The parameterised approach;
- The numerical approach.

The module ***efgtgro*** allows the estimation of a poverty elasticity or semi-elasticity with respect to growth with the different approaches mentioned above. For more details on these approaches, see:

Abdelkrim Araar, 2012. "[Expected Poverty Changes with Economic Growth and Redistribution](#)" *Cahiers de recherche* 1222, CIRPEE.

- To estimate a FGT elasticity –semi-elasticity- with respect to average income growth in a group or in an entire population;
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- The results are displayed with 6 decimals; this can be changed.

## 12.3 FGT elasticity with respect to Gini inequality (*efgtineq*).

The overall growth elasticity (INEL) of poverty, when growth comes exclusively from change in inequality within a group  $k$ , is given by:

$$INEL = \begin{cases} \frac{\phi(k)f(k,z)(\mu(k) - z)}{F(z)} / \frac{\phi(k)\mu(k)}{\mu} \frac{C(k)}{I} & \text{if } \alpha = 0 \\ \alpha \frac{\bar{P}(k,z;\alpha) - ((\mu(k) - z)/z)\bar{P}(k,z;\alpha-1)}{\bar{P}(z,\alpha)} / \frac{\phi(k)\mu(k)}{\mu} \frac{C(k)}{I} & \text{if } \alpha \geq 1 \end{cases}$$

where  $z$  is the poverty line,  $k$  is the population subgroup in which growth takes place,  $f(k, z)$  is the density function at level of income  $z$  for group  $k$ , and  $F(z)$  is the headcount.  $C(k)$  is the concentration coefficient of group  $k$  when incomes of the complement group are replaced by  $\mu(k)$ .  $I$  denotes the Gini index.

Araar, Abdelkrim and Jean-Yves Duclos, (2007), [Poverty and inequality components: a micro framework, Working Paper: 07-35](#). CIRPEE, Department of Economics, Université Laval.

Kakwani, N. (1993) "Poverty and economic growth with application to Côte d'Ivoire", Review of Income and Wealth, 39(2): 121:139.

- To estimate a FGT elasticity with respect to average income growth in a group or in an entire population;
- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

## 12.4 FGT elasticity with respect to Gini-inequality with different approaches (efgtine).

The overall Gini-inequality elasticity of poverty can be estimated by using one approach among the following list:

- The counterfactual approach;
- The marginal approach;
- The parameterised approach;
- The numerical approach.

The module **efgtine** allows the estimation of a poverty elasticity or semi-elasticity with respect to inequality with the different approaches mentioned above. For more details on these approaches, see:

Abdelkrim Araar, 2012. "[Expected Poverty Changes with Economic Growth and Redistribution](#)" *Cahiers de recherche* 1222, CIRPEE.

- To estimate a FGT elasticity –semi-elasticity- with respect to inequality;
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- The results are displayed with 6 decimals; this can be changed.

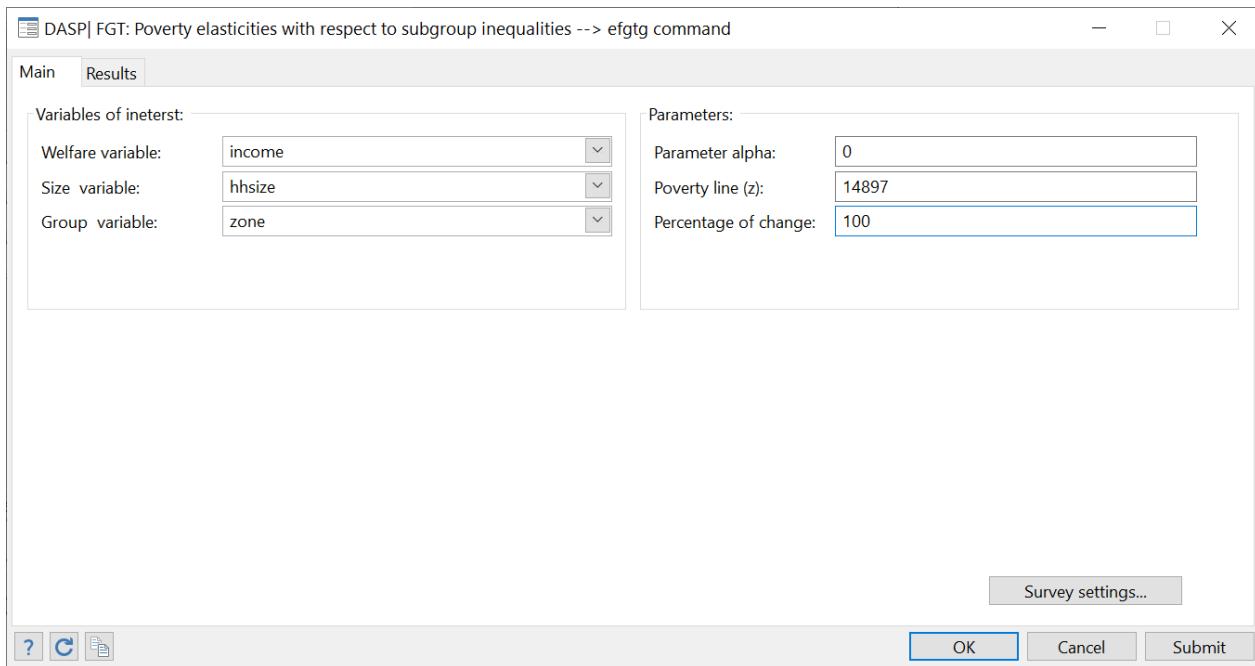
## 12.5 FGT elasticities with respect to within/between group components of inequality (efgtg).

This module estimates the marginal FGT impact and FGT elasticity with respect to within/between group components of inequality. A group variable must be provided. This module is mostly based on Araar and Duclos (2007):

Araar, Abdelkrim and Jean-Yves Duclos, (2007), [Poverty and inequality components: a micro framework](#), Working Paper: 07-35. CIRPEE, Department of Economics, Université Laval.

To open the dialog box of this module, type the command **db efgtg**.

**Figure 11: The EFGTG dialog box**



After clicking on **SUBMIT**, the following should be displayed:

```
efgtg income, hgroup(zone) hsize(hhszie) alpha(0) pline(14897) prc(1) dec(3)
```

Poverty and Inequality Indices

Indices	Estimate
FGT	<b>0.585</b>
Gini	<b>0.617</b>

Marginal Impact & Elasticities By Groups

Group	Population Share	Marginal Impact on Ineq.	Marginal Impact on Pov.	Elasticity
1: South south	0.150	0.110	0.036	0.350
2: South east	0.119	0.076	0.014	0.196
3: South west	0.194	0.138	0.077	0.588
4: North central	0.139	0.065	0.005	0.078
5: North east	0.135	0.060	0.002	0.029
6: North west	0.263	0.090	-0.008	-0.090
σ(g)	Within	0.539	0.126	0.247
σ				
γ	Between	0.051	0.051	1.045
λ	Population	1.000	0.617	0.231

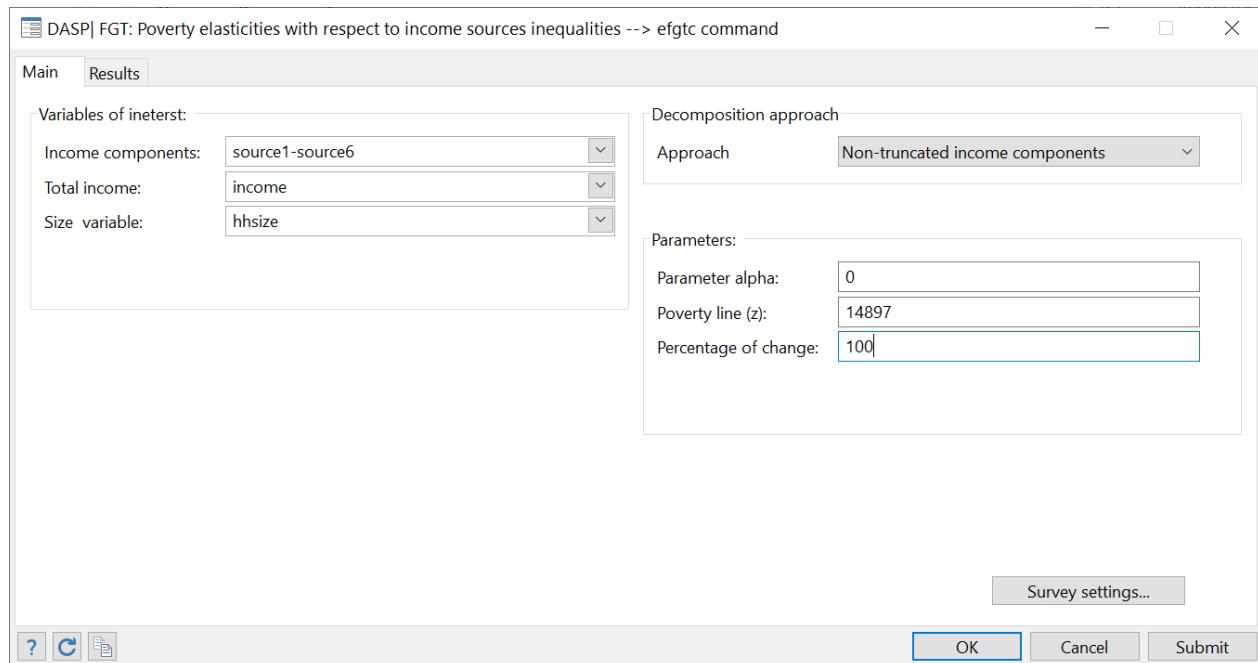
## 12.6 FGT elasticities with respect to within/between income components of inequality (efgtc).

This module estimates the marginal FGT impact and FGT elasticity with respect to within/between income components of inequality. A list of income components must be provided. This module is mostly based on Araar and Duclos (2007):

Araar, Abdelkrim and Jean-Yves Duclos, (2007), [Poverty and inequality components: a micro framework](#), Working Paper: 07-35. CIRPEE, Department of Economics, Université Laval.

To open the dialog box of this module, type the command `db efgtc`.

Figure 12: The EFGTC dialog box



After clicking on **SUBMIT**, the following should be displayed:

```
efgtc source1- source6, tot(income) hsize(hhsiz) alpha(0) pline(14897) prc(1)
```

Poverty and Inequality Indices

Indices	Estimate
FGT	<b>0.584667</b>
Gini	<b>0.616503</b>

Marginal Impacts & Elasticities of poverty with respect to the within/between inequality in income components

	Source	Income Share	Impact on Inequality	Impact on Poverty	Elasticity
$\eta(k)$	1: source1	<b>0.352966</b>	<b>0.265888</b>	<b>0.097233</b>	<b>0.385605</b>
	2: source2	<b>0.199865</b>	<b>0.063585</b>	<b>-0.032419</b>	<b>-0.537610</b>
	3: source3	<b>0.023731</b>	<b>0.012489</b>	<b>0.002508</b>	<b>0.211784</b>
	4: source4	<b>0.344093</b>	<b>0.229384</b>	<b>0.067828</b>	<b>0.311798</b>
	5: source5	<b>0.024588</b>	<b>0.013828</b>	<b>0.002247</b>	<b>0.171358</b>
	6: source6	<b>0.054758</b>	<b>0.031356</b>	<b>0.005368</b>	<b>0.180533</b>
$\lambda$	Within	.	<b>0.616503</b>	<b>0.134793</b>	<b>0.230546</b>
$\tau$	Between	.	<b>0.049948</b>	<b>0.041846</b>	<b>0.883417</b>

In case one is interested in changing some income component only among those individuals that are effectively active in some economic sectors (schemes  $\eta^*(k)$ ,  $\tau^*$  and  $\lambda^*$  in the paper mentioned above), the user should select the approach "Truncated income component".

## 13 DASP and inequality indices

Gini and concentration indices (igini)

### 13.1 Inequality indices (ineq)

The **ineq** module is used to estimate the different inequality indices and their standard errors. The user must add the index option (index\_name) to estimate the desired index.

Index name	Inequality index
<b>gini</b>	Gini index
<b>agini</b>	Absolute Gini index
<b>conc</b>	Concentration index
<b>aconc</b>	Absolute concentration index
<b>atk</b>	Atkinson index
<b>entropy</b>	Generalised entropy index
<b>covar</b>	Coefficient of variation index
<b>qr</b>	Quantile ratio index

- The user can select more than one variable of interest at a time. For example, inequality can be estimated by simultaneously using *per capita* consumption and *per capita* income.
- A group variable can be used to estimate inequality at the level of a categorial group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.

### 13.1.1 Gini and concentration indices

The Gini index is estimated as

$$\hat{I} = 1 - \frac{\hat{\xi}}{\hat{\mu}}$$

where

$$\hat{\xi} = \sum_{i=1}^n \left[ \frac{(V_i)^2 - (V_{i+1})^2}{[V_1]^2} \right] y_i \quad \text{and} \quad V_i = \sum_{h=i}^n w_h \quad \text{and} \quad y_1 \geq y_2 \geq \cdots \geq y_{n-1} \geq y_n.$$

The absolute Gini index is estimated as

$$\widehat{AI} = \hat{\mu} - \hat{\xi}$$

When the ranking variable is  $Y$ , the concentration index for the variable  $T$  is estimated as

$$\widehat{IC}_T = 1 - \frac{\hat{\xi}_T}{\hat{\mu}_T}$$

where  $\hat{\mu}_T$  is the average of variable  $T$

$$\hat{\xi}_T = \sum_{i=1}^n \left[ \frac{(V_i)^2 - (V_{i+1})^2}{[V_1]^2} \right] t_i$$

where  $V_i = \sum_{h=i}^n w_h$  and  $y_1 \geq y_2 \geq \cdots \geq y_{n-1} \geq y_n$ .

The absolute concentration index is estimated as

$$\widehat{IAC}_T = \hat{\mu}_T - \hat{\xi}_T$$

Interested users are encouraged to consider the exercises that appear in Section 23.12. The user must select a ranking variable to estimate a concentration index.

### 13.1.2 Generalised entropy index

The generalised entropy index is estimated as

$$\hat{I}(\theta) = \begin{cases} \frac{1}{\theta(\theta-1)\sum_{i=1}^n w_i} \sum_i w_i \left[ \left( \frac{y_i}{\hat{\mu}} \right)^\theta - 1 \right] & \text{if } \theta \neq 0, 1 \\ \frac{1}{\sum_{i=1}^n w_i} \sum_i w_i \log \left( \frac{\hat{\mu}}{y_i} \right) & \text{if } \theta = 0 \\ \frac{1}{\sum_{i=1}^n w_i} \sum_i \frac{w_i y_i}{\hat{\mu}} \log \left( \frac{y_i}{\hat{\mu}} \right) & \text{if } \theta = 1 \end{cases}$$

### 13.1.3 Atkinson index

The Atkinson index of inequality for the group k is denoted by  $I(\varepsilon)$ . It can be expressed as:

$$\hat{I}(\varepsilon) = \frac{\hat{\mu} - \hat{\xi}(\varepsilon)}{\hat{\mu}} \quad \text{where} \quad \hat{\mu} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

The Atkinson index of social welfare is expressed as:

$$\hat{\xi}(\varepsilon) = \begin{cases} \left[ \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i (y_i)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} & \rightarrow \text{if } \varepsilon \neq 1 \quad \text{and } \varepsilon \geq 0 \\ \exp \left[ \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i \ln(y_i) \right] & \rightarrow \varepsilon = 1 \end{cases}$$

- The user can select more than one variable of interest at a time. For example, inequality can be estimated for *per capita* consumption and *per capita* income simultaneously.

#### 13.1.4 Coefficient of variation index

The coefficient of variation index of inequality for group k is denoted by CV. It can be expressed as:

$$\widehat{CV} = \left[ \frac{\sum_{i=1}^n w_i y_i^2 / \sum_{i=1}^n w_i - \hat{\mu}^2}{\hat{\mu}^2} \right]^{\frac{1}{2}}$$

#### 13.1.5 Quantile/share ratio indices of inequality

The quantile ratio is estimated as

$$\widehat{QR}(p_1, p_2) = \frac{\widehat{Q}(p_1)}{\widehat{Q}(p_2)}$$

where  $Q(p)$  denotes a p-quantile, and  $p_1$  and  $p_2$  are percentiles.

The share ratio is estimated as

$$\widehat{SR}(p_1, p_2, p_3, p_4) = \frac{\widehat{GL}(p_2) - \widehat{GL}(p_1)}{\widehat{GL}(p_4) - \widehat{GL}(p_3)}$$

where  $GL(p)$  is the generalised Lorenz curve, and  $p_1, p_2, p_3$  and  $p_4$  are percentiles.

## 13.2 Differences between inequality indices (dineq)

The **dineq** module estimates differences between the inequality indices of two distributions.

For each distribution:

- One variable of interest should be selected.
- A ranking variable must be selected to estimate a concentration index.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.

See the Section [13.1](#) for more information on inequality indices.

### 13.3 The Araar (2009) multidimensional inequality index

The Araar (2009) the multidimensional inequality index for the  $K$  dimensions of well-being takes the following form:

$$I = \sum_{i=1}^K \phi_k [\lambda_k I_k + (1 - \lambda_k) C_k]$$

where  $\phi_k$  is the weight attributed to the dimension  $k$  (may take the same value across the dimensions or can depend on the averages of the well-being dimensions).  $I_k, C_k$  are respectively the relative –absolute–Gini and concentration indices of component  $k$ . The normative parameter  $\lambda_k$  controls the sensitivity of the index to the inter-correlation between dimensions. For more details, see:

Abdelkrim Araar, 2009. "The Hybrid Multidimensional Index of Inequality," Cahiers de recherche 0945, CIRPEE: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1496505](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496505)

## 14 DASP and polarisation indices

### 14.1 Polarisation indices (ipola)

The **ipola** module can be used to estimate the different polarisation indices and their standard errors. The user must add the index option (index\_name) to estimate the desired index.

Index name	Inequality index
der	Duclos, Esteban and Ray (2004) polarisation index
fw	Foster and Wolfson (1992) polarisation index
egr	Generalised Esteban, Gradian and Ray (1999) polarisation index
in	Inaki (2008) polarisation index

- The user can select more than one variable of interest at a time. For example, polarisation can be estimated by simultaneously using per capita consumption and per capita income.
- A group variable can be used to estimate polarisation at the level of a categorial group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.

### 14.1.1 The Duclos, Esteban and Ray (2004) polarisation index

The Duclos, Esteban and Ray (2004) polarisation index (DER) can be expressed as:

$$DER(\alpha) = \iint f(x)^{1+\alpha} f(y) |y - x| dy dx$$

where  $f(x)$  denotes the density function at  $x$ .

The discrete formula that is used to estimate this index is:

$$DER(\alpha) = \frac{\sum_{i=1}^n w_i f(y_i)^\alpha a(y_i)}{\sum_{i=1}^n w_i}$$

The normalised DER estimated by the ipola module is defined as:

$$\overline{DER}(\alpha) = \frac{DER(\alpha)}{2\mu^{(1-\alpha)}}$$

where

$$a(y_i) = \mu + y_i \left( \left( \frac{2 \sum_{j=1}^i w_j - w_i}{\sum_{i=1}^N w_i} \right) - 1 \right) - \left( \frac{2 \sum_{j=1}^{i-1} w_j y_j + w_i y_i}{\sum_{i=1}^N w_i} \right)$$

The Gaussian kernel estimator is used to estimate the density function.

Main reference

DUCLOS, J.-Y., J. ESTEBAN, AND D. RAY (2004): "Polarisation: Concepts, Measurement, Estimation," *Econometrica*, 72, 1737–1772.

### 14.1.2 The Foster and Wolfson (1992) polarisation index

The Foster and Wolfson (1992) polarisation index (FW) can be expressed as:

$$FW = 2[2[0.5 - Lorenz(p = 0.5)] - Gini] \frac{\mu}{median}$$

Main reference

FOSTER, J. AND M. WOLFSON (1992): "Polarisation and the Decline of the Middle Class: Canada and the U.S." mimeo, Vanderbilt University.

### 14.1.3 The generalised Esteban, Gradiñ and Ray (1999) polarisation index

The polarisation index proposed by Esteban and Ray (1994) is defined as:

$$P^{ER}(f, \alpha) = \sum_{j=1}^m \sum_{k=1}^m p_j^{1+\alpha} p_k |\mu_j - \mu_k|$$

where  $\mu_j$  and  $p_j$  denote the average income and the population share of group  $j$ , respectively. Parameter  $\alpha \in [1, 1.6]$  reflects society's sensitivity to polarisation. The first step of the estimation process requires defining exhaustive and mutually exclusive groups,  $\rho$ . This involves some degree of error. The polarisation index proposed by Esteban et al. (1999) takes this idea into account and is obtained after correcting the  $P^{ER}(\alpha)$  index applied to the simplified representation of the original distribution with a measurement of the grouping error. Nonetheless, when dealing with personal or spatial income distributions, there are no unanimous criteria for demarcating different groups. To address this problem, Esteban et al. (1999) follow the methodology proposed by Aghevli and Mehran (1981) and Davies and Shorrocks (1989) to find an optimal distribution partition for a given number of groups,  $\rho^*$ . This means selecting the partition that minimises the Gini index value of within-group inequality,  $Error = G(f) - G(\rho^*)$  (see Esteban et al., 1999). The polarisation index proposed by Esteban et al. (1999) is therefore given by:

$$P^{EGR}(f, \alpha, \rho^*, \beta) = \sum_{j=1}^m \sum_{k=1}^m p_j^{1+\alpha} p_k |\mu_j - \mu_k| - \beta(G(f) - G(\rho^*))$$

where  $\beta \geq 0$  is a parameter that informs the weight assigned to the error term. (In Esteban et al., 1999, the value used is  $\beta = 1$ .)

The ipola module with the index option (egr) estimates the generalised form of the Esteban et al. (1999) polarisation index. In addition to the usual variables, this routine offers the following three options:

1. The number of groups, which the user can select. Empirical studies use two or three groups. The program searches for and displays the optimal income interval for each group depending on this number. It also displays the error as a percentage, i.e.,  $\frac{G(f) - G(\rho^*)}{G(f)} * 100$ ;
2. Parameter  $\alpha$ ; and
3. Parameter  $\beta$ .

To respect the scale invariance principle, all incomes are divided by the average income i.e.,  $(\dot{\mu}_j = \mu_j / \mu)$ . In addition, the index is divided by the scalar 2 to make its interval lie between 0 and 1 when  $\alpha = 1$ .

$$\dot{P}^{EGR}(f, \alpha, \rho^*, \beta) = 0.5 \left( \sum_{j=1}^m \sum_{k=1}^m p_j^{1+\alpha} p_k |\dot{\mu}_j - \dot{\mu}_k| - \beta(G(f) - G(\rho^*)) \right)$$

#### 14.1.4 The Inaki (2008) polarisation index

Let a population be split into  $N$  groups of size  $n_i > 0$ . The density function, the mean and the population share of group  $i$  are denoted by  $f_i(x)$ ,  $\mu_i$  and  $\pi_i$ , respectively. The variable  $\mu$  is the overall mean. Therefore,  $\int f_i(x) = 1$ ,  $\sum_{i=1}^N \pi_i \mu_i = \mu$  and  $\sum_{i=1}^N \pi_i = 1$ . The following social polarisation index can be defined using Inaki (2008):

$$P(F) = \sum_{i=1}^N [P_W(i, F) + P_B(i, F)]$$

where

$$P_W(i, F) = \mu^{\alpha-1} \pi_i^{2+\alpha} \int \int f_i^{-1+\alpha}(x) f_i(y) |x - y| dy dx$$

and

$$P_B(i, F) = \mu^{\alpha-1} \pi_i^{1+\alpha} \left[ (\mu - \pi_i \mu_i) \int f_i^{-1+\alpha}(x) dx + (1 - \pi_i) \int f_i^{-1+\alpha}(x) x dx \right]$$

The **ipola** module with the index option (in) makes it possible to decompose the social polarisation index  $P(F)$  into group components. The results show:

- The estimated population share of subgroup  $i$ ,  $\pi_i$ ;
- The estimated income share of subgroup  $i$ ,  $\pi_i \mu_i / \mu$ ;
- The estimated  $P_W(i, F)$  index of subgroup  $i$ ;
- The estimated  $P_B(i, F)$  index of subgroup  $i$ ;
- The estimated  $P_W = \sum_i P_W(i, F)$  index;
- The estimated  $P_B = \sum_i P_B(i, F)$  index; and
- The estimated total index  $P_F$ .

#### Example:

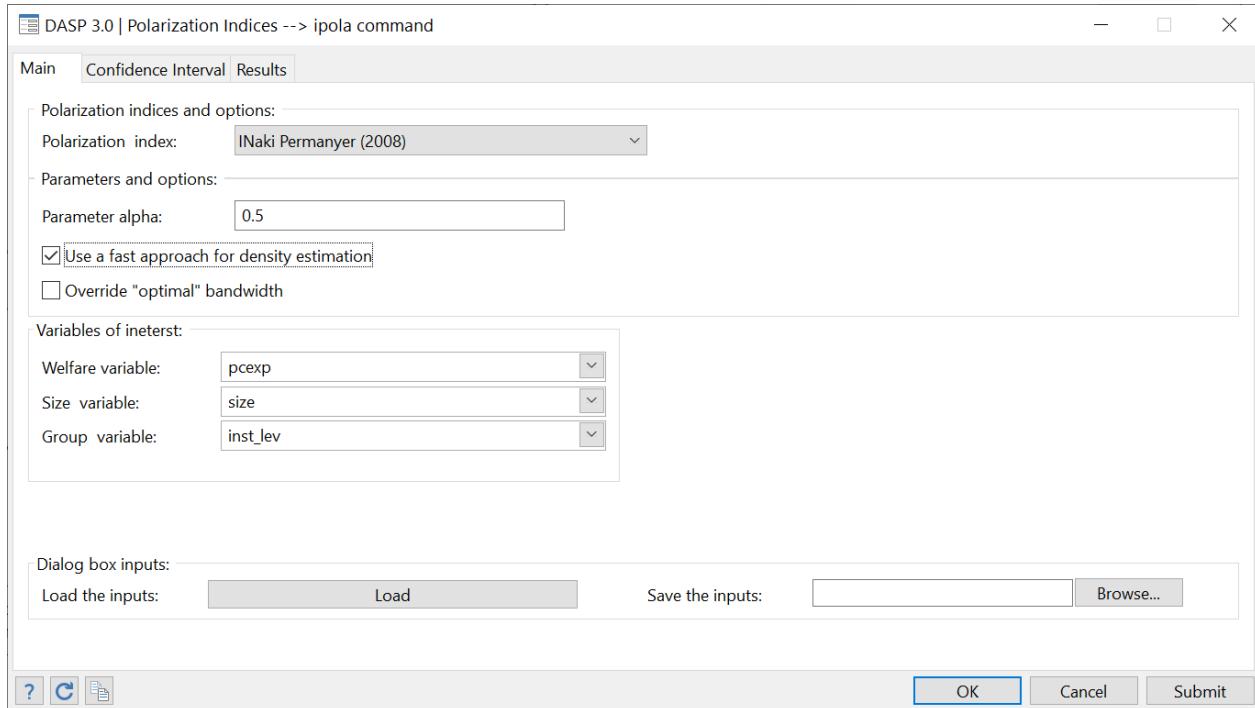
For illustrative purposes, we use a 1996 Cameroonian household survey that includes approximately 1700 households. The variables used are:

<b>STRATA</b>	The stratum in which a household lives
<b>PSU</b>	The household's primary sampling unit
<b>WEIGHT</b>	The sampling weight
<b>SIZE</b>	The household size
<b>INS_LEVEL</b>	The education level of the head of the household <ol style="list-style-type: none"> <li>1. Primary;</li> <li>2. Professional training, secondary and superior; or</li> <li>3. Not responding.</li> </ol>

We decompose the above social polarisation index by splitting the Cameroonian population into three exclusive groups according to the education level of the head of the household. We first initialise the

survey's sampling design with the dialog box **svyset**. After that, we open the dialog box by typing *db ipola*, and choose variables and parameters as in Figure 13 below.

Figure 13: The IPOLA dialog box



After clicking on **SUBMIT**, the following results appear:

```
: . : . : . : . : <== END
Decomposition of the social polarization index by population groups
Household size : size
Sampling weight : weight
Group variable : inst_lev
Parameter alpha : 0.50
Social polarization index : 0.62128225 (0.00352956)
```

Group	Population Share	Income Share	Within-Group Component	Between-Group Component
No education	0.317022 0.007636	0.216402 0.006103	0.021533 0.001343	0.197993 0.006612
Primary level	0.349053 0.007344	0.276644 0.006817	0.029164 0.001575	0.200097 0.005221
Secondary level	0.274048 0.006358	0.344870 0.007920	0.021358 0.001247	0.132201 0.003760
Tertiary level	0.059877 0.002883	0.162084 0.008456	0.000847 0.000108	0.018090 0.001199
Total	1.000000 0.000000	1.000000 0.000000	0.072901 0.000958	0.548381 0.003721

### Main references

DUCLOS, J.-Y., J. ESTEBAN, AND D. RAY (2004): "Polarisation: Concepts, Measurement, Estimation," *Econometrica*, 72, 1737–1772.

Tian Z. & all (1999) "Fast Density Estimation Using CF-kernel for Very Large Databases".  
<http://portal.acm.org/citation.cfm?id=312266>

Iñaki Permanyer, 2008. "[The Measurement of Social Polarisation in a Multi-group Context](#)," UFAE and IAE Working Papers 736.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).

## 14.2 Differences between polarisation indices (dipola)

The **dipola** module estimates differences between the polarisation indices of two distributions.

For each distribution:

- One variable of interest should be selected.
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence interval and the level of confidence can be changed.
- The results are displayed with 6 decimals; this can be changed.

# 15 DASP and decompositions

## 15.1 FGT Poverty: decomposition by population subgroups (dfgtg)

The **dfgt** module decomposes the FGT poverty index by population subgroups. This decomposition takes the form

$$\widehat{P}(z; \alpha) = \sum_{g=1}^G \widehat{\phi}(g) \widehat{P}(z; \alpha; g)$$

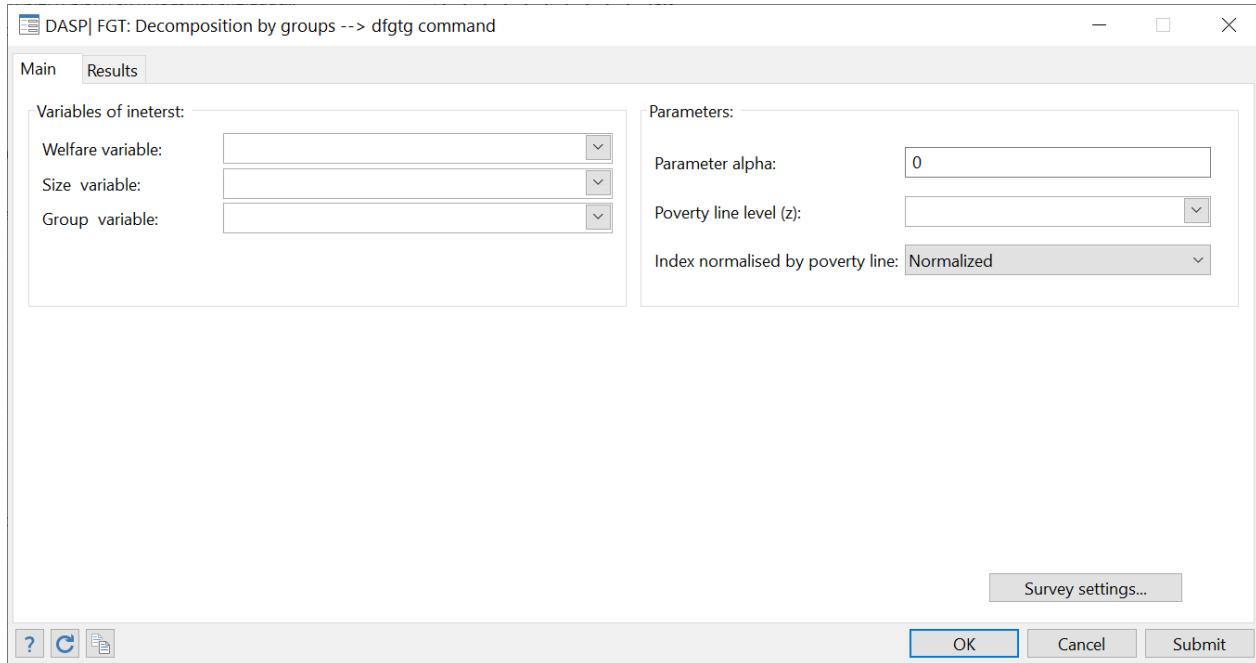
where  $G$  is the number of population subgroups. The results show:

- The estimated FGT index of subgroup  $g$ :  $\widehat{P}(z; \alpha; g)$
- The estimated population share of subgroup  $g$ :  $\widehat{\phi}(g)$
- The estimated absolute contribution of subgroup  $g$  to total poverty:  $\widehat{\phi}(g) \widehat{P}(z; \alpha; g)$
- The estimated relative contribution of subgroup  $g$  to total poverty:  $(\widehat{\phi}(g) \widehat{P}(z; \alpha; g)) / \widehat{P}(z; \alpha)$

An asymptotic standard error is provided for each of these statistics.

To open the dialog box for module **dfgtg**, type `db dfgtg` in the command window.

**Figure 14: Decomposition of the FGT index by groups**



Note that the user can save results in Excel format.

Interested users are encouraged to consider the exercises that appear in Section [23.10](#)

## 15.2 FGT Poverty: decomposition by income components using the Shapley value (dfgts)

The **dfgts** module decomposes the total alleviation of FGT poverty into a sum of the contributions generated by separate income components. Total alleviation is maximal when all individuals have an income greater than or equal to the poverty line. A negative sign on a decomposition term indicates that an income component reduces poverty. Assume that there exist  $K$  income sources and that  $s_k$  denotes income source  $k$ . The FGT index is defined as:

$$\hat{P} \left( z; \alpha; y = \sum_{k=1}^K s_k \right) = \frac{\sum_{i=1}^n w_i (1 - y/z)_+^\alpha}{\sum_{i=1}^n w_i}$$

where  $w_i$  is the weight assigned to individual  $i$  and  $n$  is sample size. The **dfgts** Stata module estimates:

- The share in total income of each income source  $k$ ;
- The absolute contribution of each source  $k$  to the value of  $(\hat{P} - 1)$ ;
- The relative contribution of each source  $k$  to the value of  $(\hat{P} - 1)$ .

Note that the **dfgts** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

Araar A and Duclos J-Y (2008), ["An algorithm for computing the Shapley Value"](#), PEP and CIRPEE. Tech.-  
Note: Novembre-2008: [http://dad.ecn.ulaval.ca/pdf\\_files/shap\\_dec\\_aj.pdf](http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf)

### **Empirical illustration with a Nigerian household survey**

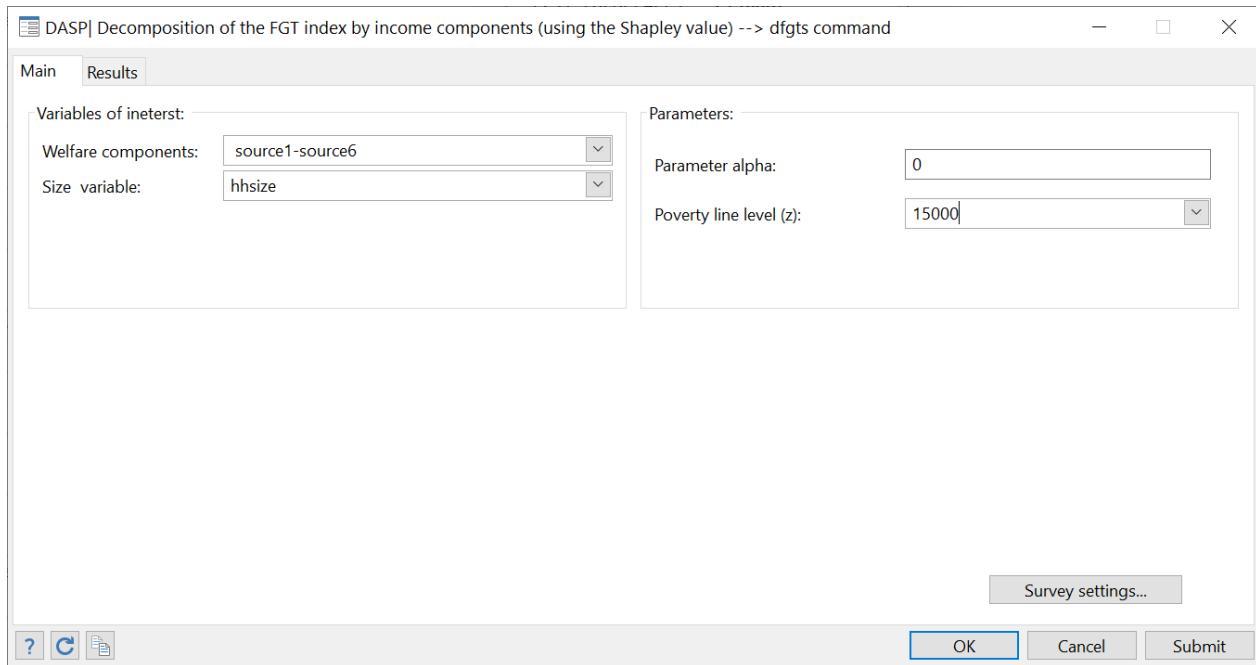
We use a survey of Nigerian households (NLSS, using 17764 observations) carried out between September 2003 and August 2004 to illustrate the use of the **dfgts** module. We use *per capita* total household income as a measure of individual living standards. Household observations are weighted by household size and sampling weights to assess poverty over all individuals. The six main income components are:

- *source\_1*: Employment income;
- *source\_2*: Agricultural income;
- *source\_3*: Fish-processing income;
- *source\_4*: Non-farm business income;
- *source\_5*: Remittances received;
- *source\_6*: All other income.

The Stata data file is saved after initialising its sampling design with the command **svyset**.

- To open the dialog box for module **dfgts**, type *db dfgts* in the command window.

**Figure 15: Decomposition of the FGT index by income components**



- Indicate the *varlist* of the six income sources.
- Indicate that the poverty line is set to 15 000 \$N.
- Set the variable HOUSEHOLD SIZE.
- Set the variable HOUSEHOLD WEIGHT.
- Click on the button SUBMIT. The following results appear:

```

. dfgts source1-source6, hsize(hysize) alpha(0) pline(15000)

Decomposition of the FGT index by income components (using the Shapley value).
Execution time :          0.61 second(s)
Parameter alpha :          0.00
Poverty line :           15000.00
FGT index :            0.584910
Household size :        hsize
Sampling weight :       weight



| Sources    | Income Share | Absolute Contribution | Relative Contribution |
|------------|--------------|-----------------------|-----------------------|
| 1: source1 | 0.352966     | -0.131599             | 0.317036              |
| 2: source2 | 0.199865     | -0.091798             | 0.221151              |
| 3: source3 | 0.023731     | -0.011535             | 0.027789              |
| 4: source4 | 0.344093     | -0.150365             | 0.362247              |
| 5: source5 | 0.024588     | -0.009299             | 0.022403              |
| 6: source6 | 0.054758     | -0.020494             | 0.049373              |
| Total      | 1.000000     | -0.415090             | 1.000000              |


```

Marginal contributions:

Source	level_1	level_2	level_3	level_4	level_5
1: source1	-0.023228	-0.022837	-0.022328	-0.021734	-0.021085
2: source2	-0.014041	-0.014698	-0.015221	-0.015637	-0.015971
3: source3	-0.001848	-0.001953	-0.001988	-0.001971	-0.001925
4: source4	-0.025759	-0.025661	-0.025484	-0.025021	-0.024543
5: source5	-0.001297	-0.001494	-0.001597	-0.001636	-0.001643
6: source6	-0.003113	-0.003378	-0.003506	-0.003536	-0.003510

Source	level_6
1: source1	-0.020387
2: source2	-0.016230
3: source3	-0.001850
4: source4	-0.023976
5: source5	-0.001631
6: source6	-0.003451

### 15.3 Alkire and Foster (2011) MD index of poverty: decomposition by population subgroups (dmdafg)

The dmdafg module decomposes the MD Alkire and Foster index of poverty index by population subgroups. This decomposition takes the form. The results show:

- The estimated Alkire and Foster index of each subgroup;
- The estimated population share of subgroup;
- The estimated absolute contribution of subgroup  $g$  to total poverty;
- The estimated relative contribution of subgroup  $g$  to total poverty.

An asymptotic standard error is provided for each of these statistics.

### 15.4 Alkire and Foster (2011: decomposition by dimensions using the Shapley value (dmdafs)

The **dmdafs** module decomposes the Alkire and Foster (2011) multidimensional poverty indices into a sum of the contributions generated by each of the poverty dimensions. It uses the Shapley characteristic function. The non-presence of a given factor –dimension- is obtained by setting the level of that dimension to its specific poverty line, thus ensuring the non-contribution of this dimension to the AF

(2011) indices. Note that the **dmdafs** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-  
Note: Novembre-2008: [http://dad.ecn.ulaval.ca/pdf\\_files/shap\\_dec\\_aj.pdf](http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf)

## 15.5 Decomposition of the variation in FGT indices into growth and redistribution components (dfgtgr)

Datt and Ravallion (1992) decompose the change in the FGT index between two periods,  $t_1$  and  $t_2$ , into growth and redistribution components as follows:

$$\underbrace{P_2 - P_1}_{\text{variation}} = \underbrace{[P(\mu^{t2}, \pi^{t1}) - P(\mu^{t1}, \pi^{t1})]}_{C_1} + \underbrace{[P(\mu^{t1}, \pi^{t2}) - P(\mu^{t1}, \pi^{t1})]}_{C_2} + R \quad /ref = 1$$

$$\underbrace{P_2 - P_1}_{\text{variation}} = \underbrace{[P(\mu^{t2}, \pi^{t2}) - P(\mu^{t1}, \pi^{t2})]}_{C_1} + \underbrace{[P(\mu^{t2}, \pi^{t2}) - P(\mu^{t2}, \pi^{t1})]}_{C_2} + R \quad /ref = 2$$

where

variation	= difference in poverty between $t_1$ and $t_2$ ;
$C_1$	= growth component;
$C_2$	= redistribution component;
R	= residual;
Ref	= period of reference.

$P(\mu^{t1}, \pi^{t1})$ : the FGT index of the first period

$P(\mu^{t1}, \pi^{t2})$ : the FGT index of the second period

$P(\mu^{t2}, \pi^{t1})$ : the FGT index of the first period when all incomes  $y_i^{t1}$  of the first period are multiplied by  $\mu^{t2}/\mu^{t1}$

$P(\mu^{t1}, \pi^{t2})$ : the FGT index of the second period when all incomes  $y_i^{t2}$  of the second period are multiplied by  $\mu^{t1}/\mu^{t2}$

The Shapley value decomposes the variation in the FGT Index between two periods,  $t_1$  and  $t_2$ , into growth and redistribution components as follows:

$$\underbrace{P_2 - P_1}_{\text{Variation}} = C_1 + C_2$$

$$C_1 = \frac{1}{2} ([P(\mu^{t2}, \pi^{t1}) - P(\mu^{t1}, \pi^{t1})] + [P(\mu^{t2}, \pi^{t2}) - P(\mu^{t1}, \pi^{t2})])$$

$$C_2 = \frac{1}{2} ([P(\mu^{t1}, \pi^{t2}) - P(\mu^{t1}, \pi^{t1})] + [P(\mu^{t2}, \pi^{t2}) - P(\mu^{t2}, \pi^{t1})])$$

## 15.6 Decomposition of change in FGT poverty by poverty and population group components –sectoral decomposition- (dfgtg2d)

Additive poverty measures, like the FGT indices, can be expressed as a sum of the poverty contributions of the various subgroups of population. Each subgroup contributes by its population share and poverty level. Thus, the change in poverty across time depends on the change in these two components. Denoting the population share of group  $k$  in period  $t$  by  $\phi^t(k)$ , the change in poverty between two periods can be expressed as (see Huppi (1991) and Duclos and Araar (2006)):

$$\begin{aligned} & P^2(z; \alpha) - P^1(z; \alpha) \\ &= \sum_k^K \underbrace{\phi^1(k)(P^2(k; z; \alpha) - P^1(k, z; \alpha))}_{\text{within-group poverty effects}} \\ &+ \sum_k^K \underbrace{P^1(k, z; \alpha)(\phi^2(k) - \phi^1(k))}_{\text{demographic or sectoral effects}} \\ &+ \sum_k^K \underbrace{((P^2(k; z; \alpha) - P^1(k, z; \alpha))(\phi^2(k) - \phi^1(k)))}_{\text{interaction term}} \end{aligned}$$

This decomposition uses the initial period as the one. If the reference period is the final, the decomposition takes the form:

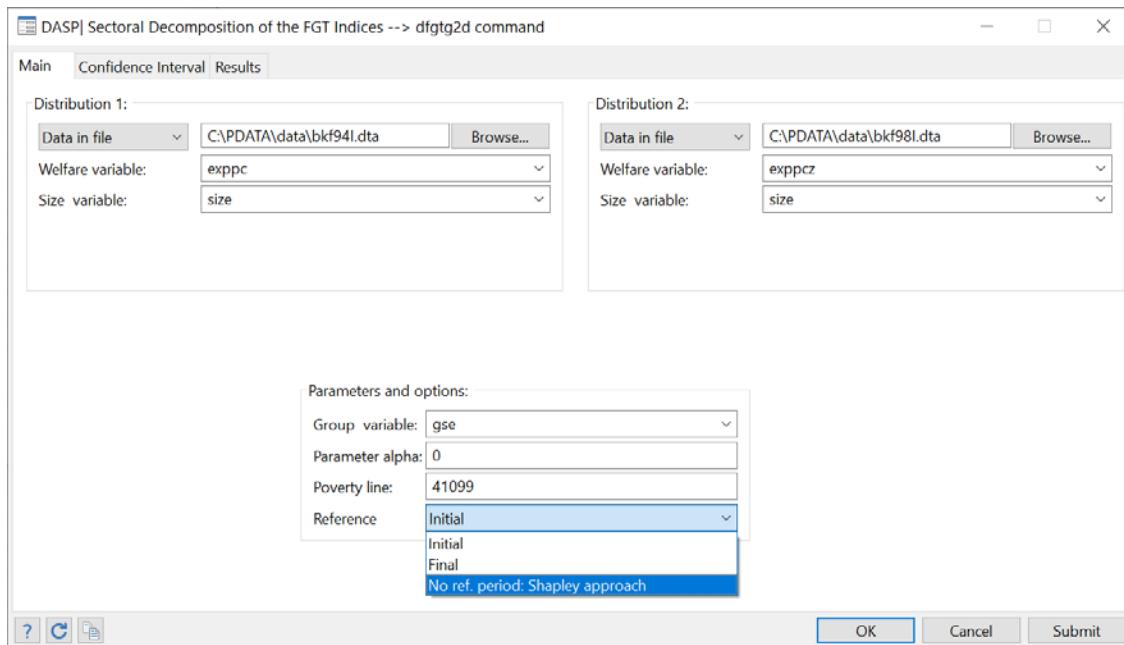
$$\begin{aligned} & P^2(z; \alpha) - P^1(z; \alpha) \\ &= \sum_k^K \underbrace{\phi^2(k)(P^2(k; z; \alpha) - P^1(k, z; \alpha))}_{\text{within-group poverty effects}} \\ &+ \sum_k^K \underbrace{P^2(k, z; \alpha)(\phi^2(k) - \phi^1(k))}_{\text{demographic or sectoral effects}} \\ &+ \sum_k^K \underbrace{((P^2(k; z; \alpha) - P^1(k, z; \alpha))(\phi^1(k) - \phi^2(k)))}_{\text{interaction term}} \end{aligned}$$

To remove the arbitrariness in selecting the reference period, we can use the Shapley decomposition approach, finding:

$$\begin{aligned}
 & P^2(z; \alpha) - P^1(z; \alpha) \\
 &= \sum_k^K \underbrace{\bar{\phi}(k)(P^2(k; z; \alpha) - P^1(k, z; \alpha))}_{\text{within-group poverty effects}} \\
 &+ \sum_k^K \underbrace{\bar{P}(k, z; \alpha)(\phi^2(k) - \phi^1(k))}_{\text{demographic or sectoral effects}}
 \end{aligned}$$

where  $\bar{\phi}(k)$  is the average population share =  $0.5(\phi^1(k) + \phi^2(k))$  and  $\bar{P}(k, z; \alpha) = 0.5(P^1(k, z; \alpha) + P^2(k, z; \alpha))$ . The DASP module dfgtg2d performs this sectoral decomposition, and this by selecting the reference period of the Shapley approach (see the following dialog box):

**Figure 16: Sectoral decomposition of FGT**



```
dfgtg2d exppc exppcz, file1(C:\PDATA\data\bkf94I.dta) hsize1(size) file2(C:\PDATA\data\bkf98I.dta) hsize2(size) alpha(0) pline(41099) hgroup(gse) ref(0)
```

Decomposition of the FGT index by groups  
 Group variable : gse  
 Parameter alpha : 0.00

Population shares and FGT indices

Group	Initial Pop. share	Initial FGT index	Final Pop. share	Final FGT index	Difference in FGT index
Wage-earner (public sector)	0.042971 0.003798	0.022406 0.012599	0.041403 0.003927	0.059094 0.023396	0.036688 0.026573
Wage-earner (private sector)	0.026598	0.067271	0.029035	0.111283	0.044812
Artisan or trader	0.002164 0.062640 0.004288	0.024093 0.097548 0.014712	0.002624 0.055795 0.004666	0.023087 0.126776 0.018202	0.033369 0.029228 0.023404
Other type of earner	0.006658 0.001308	0.194481 0.060817	0.005689 0.000923	0.293404 0.089688	0.098923 0.108357
Crop farmer	0.104482 0.014896	0.500707 0.034911	0.167806 0.014125	0.424391 0.024457	-0.076316 0.042625
Subsistence farmer	0.680885 0.016403	0.514999 0.021132	0.653552 0.015083	0.533956 0.011572	0.018957 0.024093
Inactive	0.075856 0.004839	0.414986 0.035336	0.046719 0.003354	0.386852 0.032348	-0.028134 0.047901
Population	1.000000 0.000000	0.444565 0.016124	1.000000 0.000000	0.452677 0.010927	0.008113 0.019477

Decomposition components

Group	Poverty Component	Population Component	Interaction Component
Wage-earner (public sector)	0.001548 0.001117	-0.000064 0.001931	0.000000 0.000000
Wage-earner (private sector)	0.001224	0.000218	0.000000
Artisan or trader	0.000938 0.001731	0.001222 -0.000768	0.000000 0.000000
Other type of earner	0.000618 0.000700	-0.000234 0.000930	0.000000 0.000000
Crop farmer	-0.010387	0.029328	0.000000
Subsistence farmer	0.005992 0.012648	0.013963 -0.014336	0.000000 0.000000
Inactive	0.016127 -0.001724	0.018726 -0.011681	0.000000 0.000000
Population	0.005650 ==	0.002463 ==	0.000000 ==

## 15.7 Decomposition of FGT poverty by transient and chronic poverty components (dtcpov)

This decomposes total poverty across time into transient and chronic components.

### The Jalan and Ravallion (1998) approach

Let  $y_i^t$  be the income of individual i in period t and  $\mu_i$  be average income over the T periods for that same individual i,  $i=1,\dots,N$ . Total poverty is defined as:

$$TP(\alpha, z) = \frac{\sum_{t=1}^T \sum_{i=1}^N w_i (z - y_i^t)_+^\alpha}{T \sum_{i=1}^N w_i}$$

The chronic poverty component is then defined as:

$$CPC(\alpha, z) = \frac{\sum_{i=1}^N w_i (z - \mu_i)_+^\alpha}{\sum_{i=1}^N w_i}$$

Transient poverty equals:

$$TPC(\alpha, z) = TP(\alpha, z) - CPC(\alpha, z)$$

### Duclos, Araar and Giles (2006) approach

Let  $y_i^t$  be the income of individual  $i$  in period  $t$  and  $\mu_i$  be average income over the  $T$  periods for individual  $i$ . Let  $\Gamma(\alpha, z)$  be the "equally-distributed-equivalent" (EDE) poverty gap such that:

$$\Gamma(\alpha, z) = [TP(\alpha, z)]^{1/\alpha}$$

Transient poverty is then defined as

$$TPC(\alpha, z) = \frac{\sum_{i=1}^N w_i \theta_i(\alpha, z)}{\sum_{i=1}^N w_i}$$

where  $\theta_i = \gamma_i(\alpha, z) - \gamma_i(1, z)$  and

$$\gamma_i(\alpha, z) = \left( \frac{(\sum_{i=t}^T (z - y_i^t)^{\alpha}_+)}{T} \right)^{\frac{1}{\alpha}},$$

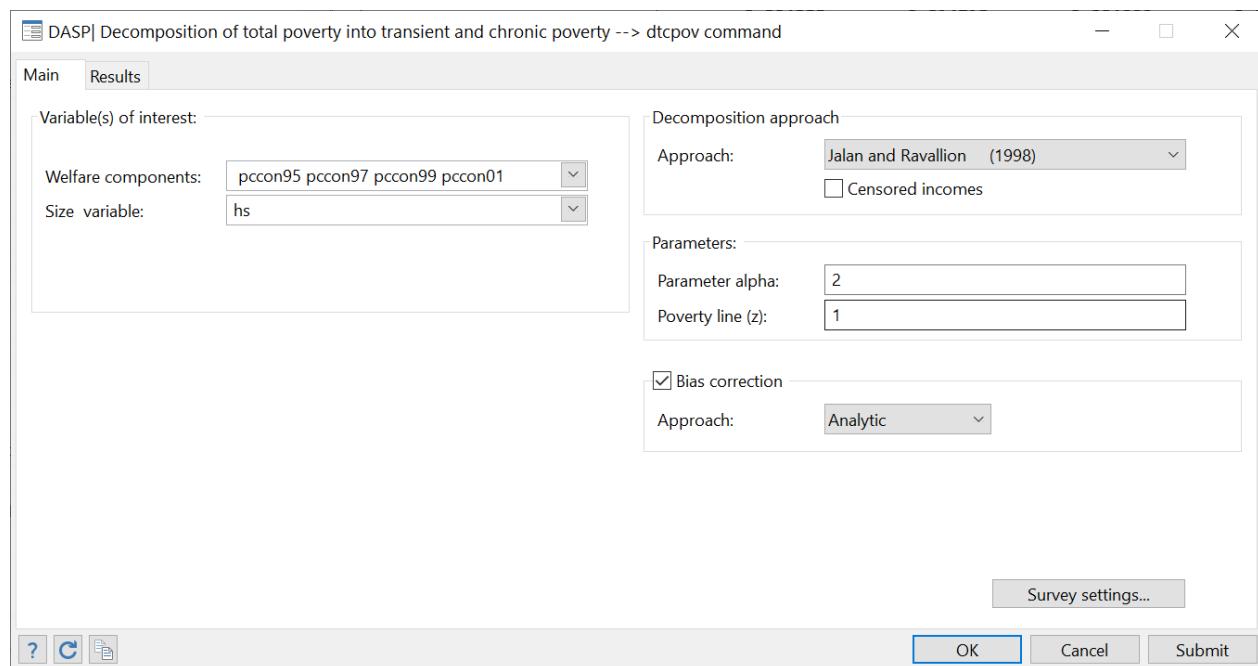
and chronic poverty is given by

$$CPC(\alpha, z) = \Gamma(\alpha, z) - TPC(\alpha, z).$$

Note that the number of periods available for this type of exercise is generally small. Because of this, a bias-correction is typically useful, using either an analytical/asymptotic or bootstrap approach.

To open the dialog box for module **dtcpov**, type `db dtcpov` in the command window.

**Figure 17: Decomposition of poverty into transient and chronic components**



- The user can select more than one variable of interest simultaneously, where each variable represents income for one period.
- The user can select one of the two approaches presented above.
- Small-T-bias-corrections can be applied, using either an analytical/asymptotic or a bootstrap approach.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided, and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

## References

Jalan Jyotsna, and Martin Ravallion. (1998) "[Transient Poverty in Postreform Rural China](#)" Journal of Comparative Economics, 26(2), pp. 338:57.

Jean-Yves Duclos & Abdelkrim Araar & John Giles, 2006. "[Chronic and Transient Poverty: Measurement and Estimation, with Evidence from China](#)," Working Paper 0611, CIRPEE.

## 15.8 Inequality: decomposition by income sources (diginis)

### Analytical approach

The **diginis** module decomposes the (usual) relative or the absolute Gini index by income sources. The three available approaches are:

- Rao's approach (1969)
- Lerman and Yitzhaki's approach (1985)
- Araar's approach (2006)

## References

Lerman, R. I., and S. Yitzhaki. "Income Inequality Effects by Income Source: A New Approach and Applications to the United States." Review of Economics and Statistics 67 (1985): 151-56.

Araar Abdelkrim (2006). On the Decomposition of the Gini Coefficient: an Exact Approach, with an Illustration Using Cameroonian Data, Working paper 02-06, CIRPEE.

### Shapley approach

The **dsineqs** module decomposes inequality indices into a sum of the contributions generated by separate income components. The **dsineqs** Stata module estimates:

- The share in total income of each income source  $k$ ;
- The absolute contribution of each source  $k$  to the Gini index;
- The relative contribution of each source  $k$  to the Gini index;

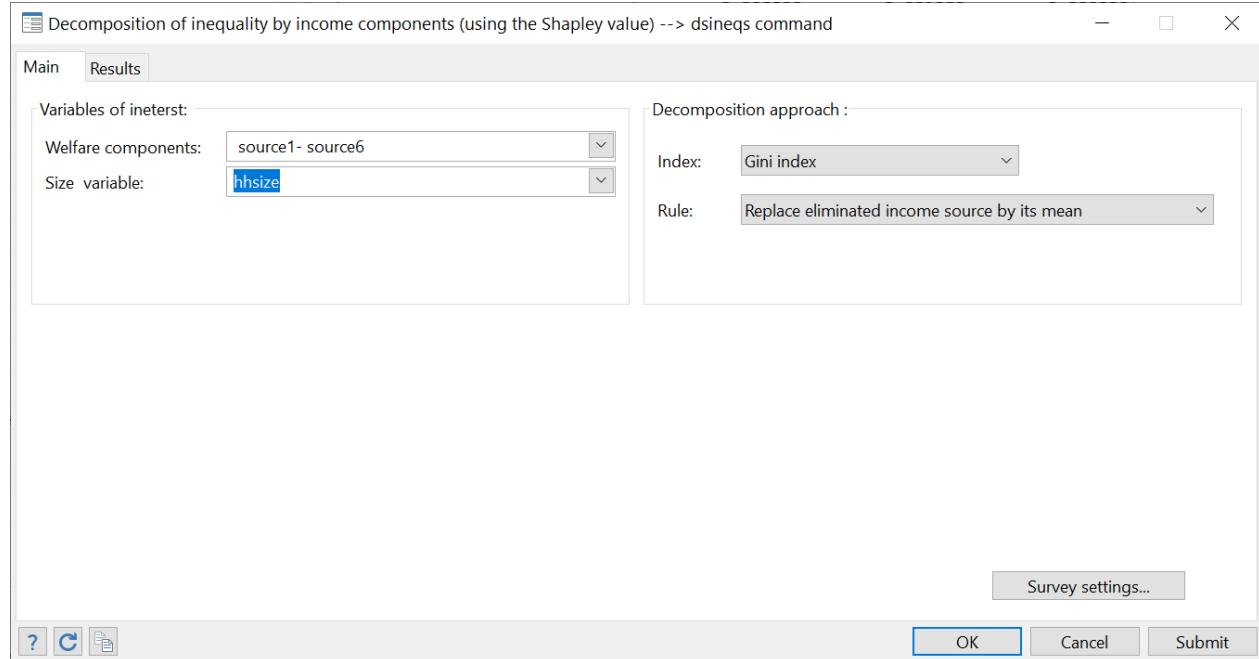
For the Shapley decomposition, the rule that is used to estimate the inequality index for a subset of components is by suppressing the inequality generated by the complement subset of components. For

this, we generate a counterfactual vector of income that equals the sum of the components of the subset plus the average of the complement subset. Note that the **dsineqs** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-  
Note: Novembre-2008: : [http://dad.ecn.ulaval.ca/pdf\\_files/shap\\_dec\\_aj.pdf](http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf)

To open the dialog box for module **dsginis**, type **db dsginis** in the command window.

**Figure 18: Decomposition of the Gini index by income sources (Shapley approach)**



## 15.9 Regression-based decomposition of inequality by income sources

A useful approach to show the contribution of income covariates to total inequality is by decomposing the latter by the predicted contributions of covariates. Formally, denote total income by  $y$  and the set of covariates by  $X = \{x_1, x_2, \dots, x_K\}$ . Using a linear model specification, we have:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k + \dots + \hat{\beta}_K x_K + \hat{\varepsilon}$$

where  $\hat{\beta}_0$  and  $\hat{\varepsilon}$  denote respectively the estimated constant term and the residual.

Two approaches for the decomposition of total inequality by income sources are used:

1. The Shapley approach: This approach is based on the expected marginal contribution of income sources to total inequality.
2. The Analytical approach: This approach is based on algebraic developments that express total inequality as a sum of inequality contributions of income sources.

With the Shapley approach:

- The user can select among the following relative inequality indices;
  - Gini index
  - Atkinson index
  - Generalised entropy index
  - Coefficient variation index
- The user can select among the following model specifications.
  - Linear :  $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_K x_K + \hat{\varepsilon}$
  - Semi Log Linear :  $\log(y) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_K x_K + \hat{\varepsilon}$

With the Analytical approach:

- The user can select among the following relative inequality indices;
  - Gini index
  - Squared coefficient variation index
- The model specification is linear.

### **Decomposing total inequality with the analytical approach**

Total income equals  $y = s_0 + s_1 + s_2 + \cdots + s_K + s_R$  where  $s_0$  is the estimated constant,  $s_k = \hat{\beta}_k X_k$  and  $s_R$  is the estimated residual. As reported by Wang 2004, relative inequality indices are not defined when the average of the variable of interest equals zero (the case of the residual). Also, inequality indices equal zero when the variable of interest is a constant (the case of the estimated constant). To deal with these two problems, Wang (2004) proposes the following basic rules:

Let  $\hat{y} = s_0 + s_1 + s_2 + \cdots + s_K$  and  $\tilde{y} = s_1 + s_2 + \cdots + s_K$ , then:  $I(y) = cs_0 + I(\tilde{y}) + cs_r$

The contribution of the constant:  $cs_0 = I(y) - I(\hat{y})$

The contribution of the residual:  $cs_r = I(\hat{y}) - I(\tilde{y})$

#### The Gini index

Using Rao 1969's approach, the relative Gini index can be decomposed as follows:

$$I(\tilde{y}) = \frac{\mu_k}{\mu_{\tilde{y}}} \tilde{C}_k$$

where  $\mu_{\tilde{y}}$  is the average of  $\tilde{y}$  and  $\tilde{C}_k$  is the coefficient of concentration of  $s_k$  when  $\tilde{y}$  is the ranking variable.

#### The Squared coefficient of variation index

As shown by Shorrocks 1982, the squared coefficient of variation index can be decomposed as:

$$I(\tilde{y}) = \sum_{k=1}^K \frac{\text{Cov}(\tilde{y}, s_k)}{\mu_{\tilde{y}}^2}$$

### Shapley decompositions

The Shapley approach is built around the expected marginal contribution of a component. The user can select among two methods to define the impact of missing a given component.

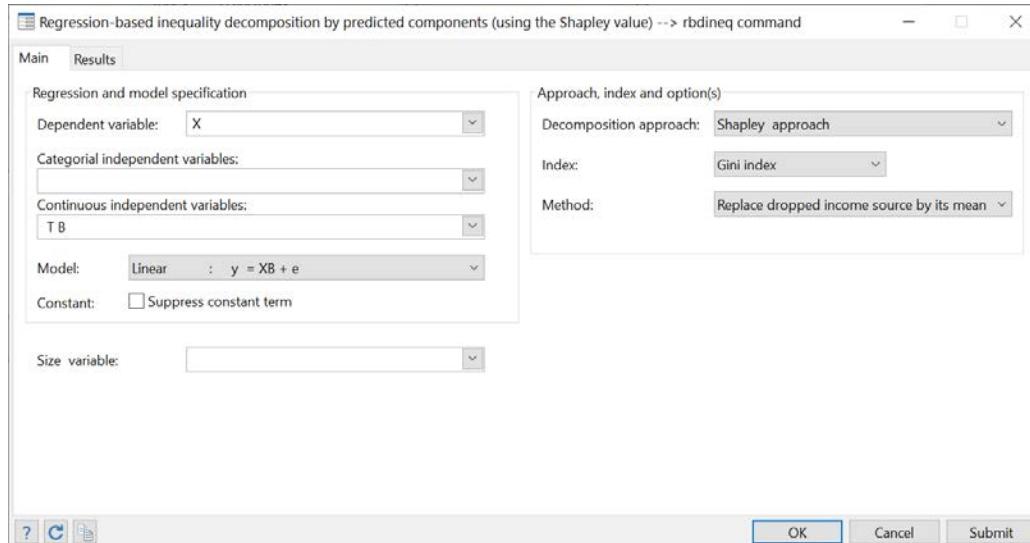
- With option: method(mean), when a component is missing from a given set of components, it is replaced by its mean.
- With option: method(zero), when a component is missing from a given set of components, it is replaced by zero.

As indicated above, we cannot estimate relative inequality for the residual component.

- For the linear model, the decomposition takes the following form:  $I(y) = I(\hat{y}) + cs_r$ , where the contribution of the residual is  $cs_r = I(y) - I(\hat{y})$ .
- For the Semi-log linear model, the Shapley decomposition is applied to all components including the constant and the residual.

With the Shapley approach, the user can use the log-linear specification. However, the user must indicate the income variable and not the log of that variable (DASP automatically runs the regression with  $\log(y)$  as the dependent variable).

### Example 1



. rbdineq X, indcon(T B) dregres(1)  
 (sum of wgt is 1,000)

Source	SS	df	MS	Number of obs	=	1,000
Model	5.0832e+11	2	2.5416e+11	F(2, 997)	=	6409.95
Residual	3.9532e+10	997	39650537.2	Prob > F	=	0.0000
Total	5.4785e+11	999	548395630	R-squared	=	0.9278
				Adj R-squared	=	0.9277
				Root MSE	=	6296.9

_000000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
T	2.538509	.0241632	105.06	0.000	2.491093 2.585926
B	-.8753363	.0464088	-18.86	0.000	-.9664065 -.7842661
_cons	11054.16	291.9053	37.87	0.000	10481.34 11626.98

Required number of computations :8

Regression-based inequality decomposition by predicted income components(using the Shapley value).  
 Execution time : 0.12 second(s)  
 Inequality index : Gini index  
 Estimated inequality : 0.508456

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_cons	0.599814	0.000000	0.000000
2: _p_T	0.572300	0.335637	0.660111
3: _p_B	-0.172114	0.090910	0.178797
4: _p_resi	0.000000	0.081909	0.161093
Total	0.000000	0.508456	1.000000

---

Marginal contributions:

---

Source	level_1	level_2	level_3
1: _p_cons	0.000000	0.000000	0.000000
2: _p_T	0.117735	0.111879	0.106023
3: _p_B	0.036160	0.030303	0.024447

## Example 2

```
. rbdineq X, indcon(T B) index(ge) theta(0.7) model(semilog) dregres(1)
Warning-> 115 OBS are omitted:
- Dependent variable should not be <= 0 with a semilog specification.
(sum of wgt is 885)
```

Source	SS	df	MS	Number of obs	=	885
Model	504.915121	2	252.45756	F(2, 882)	=	263.42
Residual	845.30956	882	.958400862	Prob > F	=	0.0000
Total	1350.22468	884	1.52740348	R-squared	=	0.3739
				Adj R-squared	=	0.3725
				Root MSE	=	.97898

_log_dep	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
T	.0000537	3.78e-06	14.19	0.000	.0000462 .0000611
B	-.0001286	8.35e-06	-15.49	0.000	-.000145 -.0001122
_cons	9.586489	.0465104	206.11	0.000	9.495205 9.677773

Required number of computations :16

```
Regression-based inequality decomposition by predicted income components(using the Shapley value).
Execution time      : 0.10      second(s)
Inequality index    : Generalized entropy index
Estimated inequality : 0.368465
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_cons	.	0.000000	0.000000
2: _p_T	.	2.052495	5.570387
3: _p_B	.	0.035821	0.097217
4: _p_resi	.	-1.719851	-4.667604
Total	.	0.368465	1.000000

Marginal contributions:

Source	level_1	level_2	level_3	level_4
1: _p_cons	0.000000	0.000000	0.000000	0.000000
2: _p_T	0.993487	0.668262	0.350511	0.040235
3: _p_B	0.021246	0.008869	0.002367	0.004139
4: _p_resi	0.056232	-0.272881	-0.594519	-0.908683

With this specification, we have  $y = \text{Exp}(s_0 + s_1 + s_2 + \dots + s_K + s_R)$ . Then:

- We cannot estimate the income share (no linear form);
- The contribution of the constant is nil.  $y = \text{Exp}(s_0) \cdot \prod_{k=1}^K \text{Exp}(s_k) \cdot \text{Exp}(s_R)$ . Adding a constant will have not any impact.

### Example 3

Regression-based inequality decomposition by predicted components (using the Shapley value) --> rbdineq command

Main Results

Regression and model specification

Dependent variable: X

Categorical independent variables:

Continuous independent variables: T B

Model: Linear :  $y = XB + e$

Constant:  Suppress constant term

Approach, index and option(s)

Decomposition approach: Analytical approach

Index: Squared coefficient of variation

Size variable:

?

OK Cancel Submit

```
. rbdineq X, indcon(T B) index(scvar) appr(analytic) noconstant dregres(1)
(sum of wgt is 1,000)
```

Source	SS	df	MS	Number of obs	=	1,000
Model	7.9109e+11	2	3.9555e+11	F(2, 998)	=	4095.28
Residual	9.6393e+10	998	96585988.5	Prob > F	=	0.0000
Total	8.8749e+11	1,000	887486572	R-squared	=	0.8914
				Adj R-squared	=	0.8912
				Root MSE	=	9827.8

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
T	2.965586	.0333531	88.91	0.000	2.900135 3.031036
B	.2660008	.0550795	4.83	0.000	.157916 .3740857

Regression-based inequality decomposition by predicted income components.

Execution time : 0.04 second(s)

Inequality index : Squared coefficient of variation index

Estimated inequality : 1.613026

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_T	0.668583	1.637250	1.015018
2: _p_B	0.052303	-0.028919	-0.017928
3: _p_resi	0.000000	-1.869025	-1.158707
Total	0.000000	1.613026	1.000000

### 15.10 Gini index: decomposition by population subgroups (diginig)

The **diginig** module decomposes the (usual) relative or the absolute Gini index by population subgroups. Let there be  $G$  population subgroups. We wish to determine the contribution of every one of those subgroups to total population inequality. The Gini index can be decomposed as follows:

$$I = \underbrace{\sum_{g=1}^G \phi_g \varphi_g I_g}_{\text{Between}} + \underbrace{\bar{I}}_{\text{Within}} + \underbrace{R}_{\text{Overlap}}$$

where

$\phi_g$  the population share of group g.

$\varphi_g$  the income share of group g.

$\bar{I}$  between-group inequality (when each individual is assigned the average income of his group).

R The residue implied by group income overlap.

## 15.11 Generalised entropy indices of inequality: decomposition by population subgroups (dentropyg)

The Generalised Entropy indices of inequality can be decomposed as follows:

$$\hat{I}(\theta) = \sum_{k=1}^K \hat{\phi}(k) \left( \frac{\hat{\mu}(k)}{\hat{\mu}} \right)^\theta \cdot \hat{I}(k; \theta) + \bar{I}(\theta)$$

where:

$\phi(k)$  is the proportion of the population found in subgroup k.

$\mu(k)$  is the mean income of group k.

$I(k; \theta)$  is inequality within group k.

$\bar{I}(\theta)$  is population inequality if each individual in subgroup k is given the mean income of subgroup k,  $\mu(k)$ .

## 15.12 Polarisation: decomposition of the DER index by population groups (dpolag)

As proposed by Araar (2008), the Duclos, Esteban and Ray index can be decomposed as follows:

$$P\bar{I} = \underbrace{\sum_g \varphi_g^{1+\alpha} \psi_g^{1-\alpha} R_g P_g}_{\text{Within}} + \underbrace{\tilde{P}}_{\text{Between}},$$

where

$$R_g = \frac{\int a_g(x)\pi_g(x)f(x)^{1+\alpha}dx}{\varphi_g \int a_g(x)f_g(x)^{1+\alpha}dx},$$

- $\varphi_g$  and  $\psi_g$  are respectively the population and income shares of group  $g$ .
- $\pi_g(x)$  denotes the local proportion of individuals belonging to group  $g$  and having income  $x$ ;
- $\tilde{P}$  is the DER polarisation index when the within-group polarisation or inequality is ignored.

The **dpolas** module decomposes the DER index by population subgroups.

Reference(s)

Abdelkrim Araar, 2008. "[On the Decomposition of Polarisation Indices: Illustrations with Chinese and Nigerian Household Surveys](#)," *Cahiers de recherche* 0806, CIRPEE.

### 15.13 Polarisation: decomposition of the DER index by income sources (dpolas)

As proposed by Araar (2008), the Duclos, Esteban and Ray index can be decomposed as follows:

$$\mathbb{P} = \sum_k \psi_k CP_k$$

where  $CP_k = \frac{\int f(x)^{1+\alpha} a_k(x) dx}{\psi_k^\alpha \mu_k^{\alpha-1}}$  and  $\psi_k$  are respectively the pseudo-concentration index and income share of income source  $k$ . The **dpolas** module decomposes the DER index by income sources.

Reference(s)

Abdelkrim Araar, 2008. "[On the Decomposition of Polarisation Indices: Illustrations with Chinese and Nigerian Household Surveys](#)," *Cahiers de recherche* 0806, CIRPEE.

## 16 DASP and curves.

### 16.1 FGT CURVES (cfgt)

FGT curves are useful distributive tools that can *inter alia* be used to:

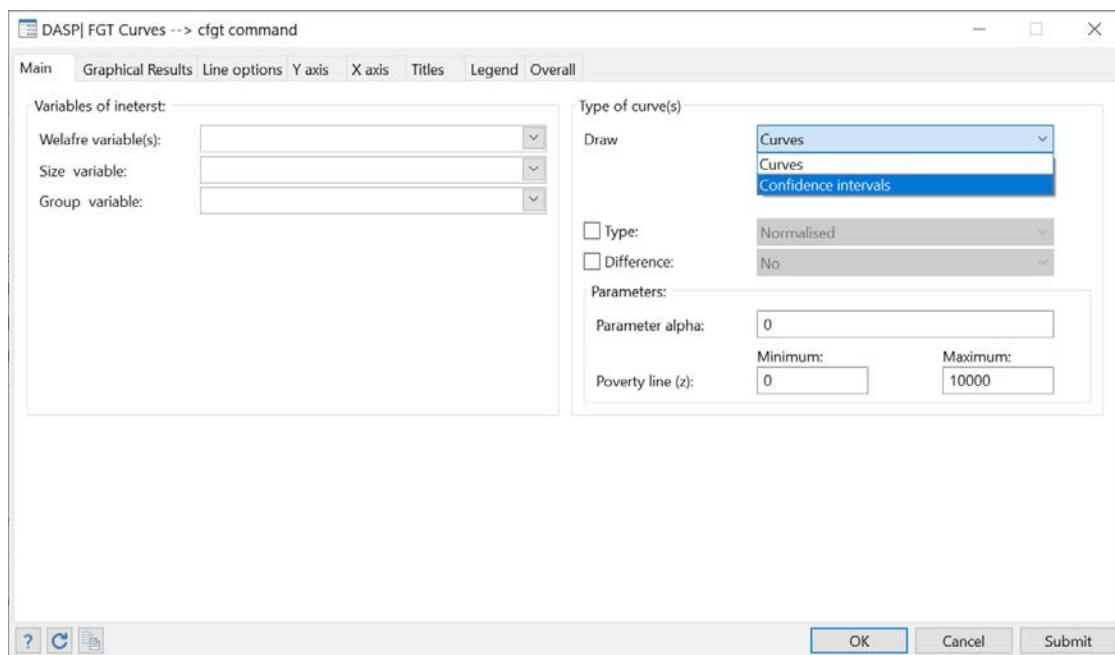
1. Show how the level of poverty varies with different poverty lines;
2. Test for poverty dominance between two distributions;
3. Test pro-poor growth conditions.

FGT curves are also called primal dominance curves. The **cfgt** module draws such curves easily. The module can:

- draw more than one FGT curve simultaneously whenever more than one variable of interest is selected.
- can draw confidence intervals instead of curves.
- draw FGT curves for different population subgroups whenever a group variable is selected.
- draw FGT curves that are not normalised by the poverty lines.
- draw differences between FGT curves.
- list or save the coordinates of the curves.
- save the graphs in different formats:
  - \*.gph : Stata format;
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **cfgt**, type the command **db dfgt** in the command window.

Figure 19: FGT curves dialog box



Interested users are encouraged to consider the exercises that appear in Section 23.7 & 23.8.

## 16.2 Difference between FGT CURVES with confidence interval (dicfgt)

The **dicfgt** module draws differences between FGT curves and their associated confidence interval by taking into account sampling design. The module can:

- draw differences between FGT curves and two-sided, lower-bounded or upper-bounded confidence intervals around these differences.
- can draw confidence intervals instead of curves.
- normalize or not the FGT curves by the poverty lines.
- list or save the coordinates of the differences between the curves as well as the confidence intervals.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section 23.8.

## 16.3 Lorenz and concentration CURVES (clorenz)

Lorenz and concentration curves are useful distributive tools that can *inter alia* be used to:

1. show the level of inequality.
2. test for inequality dominance between two distributions.
3. test for welfare dominance between two distributions.
4. test for progressivity.

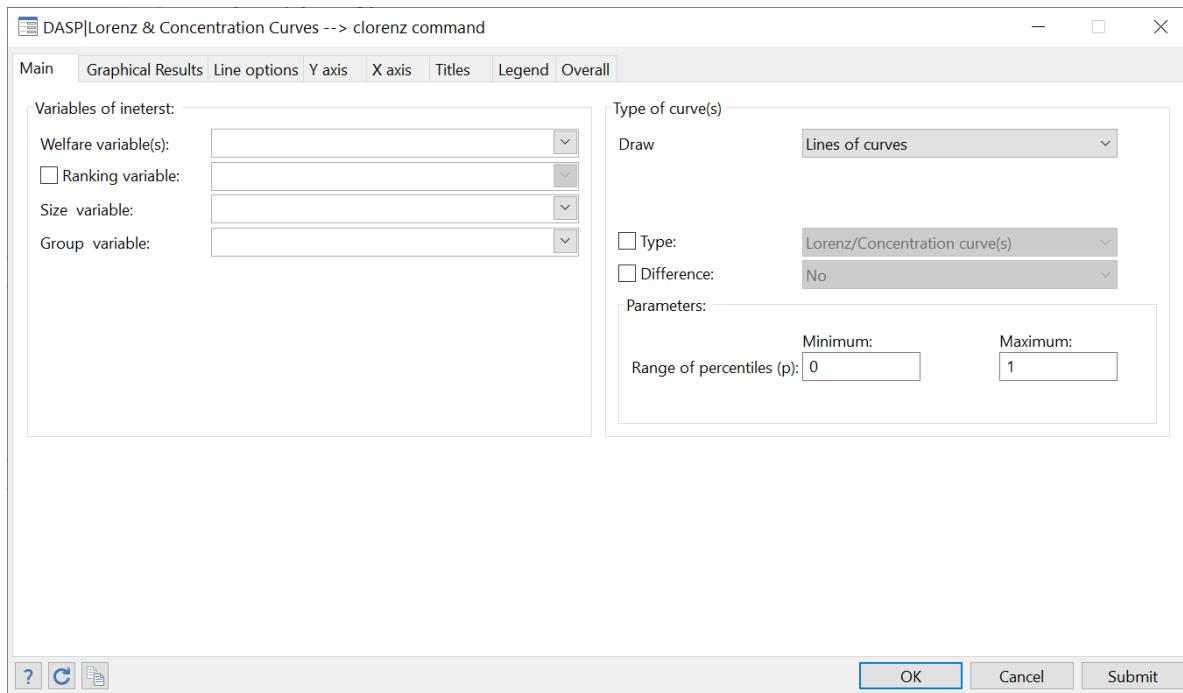
The **clorenz** module draws Lorenz and concentration curves simultaneously. The module can:

- draw more than one Lorenz or concentration curve simultaneously whenever more than one variable of interest is selected.
- can draw confidence intervals instead of curves.
- draw more than one generalised or absolute Lorenz or concentration curve simultaneously whenever more than one variable of interest is selected.
- draw more than one deficit share curve.
- draw Lorenz and concentration curves for different population subgroups whenever a group variable is selected.
- draw differences between Lorenz and concentration curves.

- list or save the coordinates of the curves.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **clorenz**, type the command `db clorenz` in the command window.

**Figure 20: Lorenz and concentration curves**



Interested users are encouraged to consider the exercises that appear in Section [23.11](#).

## 16.4 Differences between Lorenz/concentration curves with confidence interval (diclorenz)

The **diclorenz** module draws differences between Lorenz/concentration curves and their associated confidence intervals by taking sampling design into account. The module can:

- draw differences between Lorenz/concentration curves and associated two-sided, lower-bounded or upper-bounded confidence intervals.
- list or save the coordinates of the differences and their confidence intervals.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.

- Many graphical options are available to change the appearance of the graphs.

## 16.5 Poverty curves (cpoverty)

The **cpoverty** module draws the poverty gap or the cumulative poverty gap curves.

- The poverty gap at a percentile  $p$  is:  $G(p; z) = (z - Q(p))_+$
- The cumulative poverty gap at a percentile  $p$ , noted by  $CPG(p; z)$ , is given by:

$$CPG(p; z) = \frac{\sum_{i=1}^n w_i (z - y_i)_+ I(y_i \leq Q(p))}{\sum_{i=1}^n w_i}$$

The module can thus:

- draw more than one poverty gap or cumulative poverty gap curves simultaneously whenever more than one variable of interest is selected.
- draw poverty gap or cumulative poverty gap curves for different population subgroups whenever a group variable is selected.
- draw differences between poverty gap or cumulative poverty gap curves.
- list or save the coordinates of the curves.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

## 16.6 Consumption dominance curves (cdomc)

Consumption dominance curves are useful tools for studying the impact of indirect tax fiscal reforms on poverty. The  $j^{th}$  Commodity or Component dominance (C-Dominance for short) curve is defined as follows:

$$CD^j(z, s) = \begin{cases} \frac{\sum_{i=1}^n w_i (z - y_i)_+^{s-2} y_i^j}{\sum_{i=1}^n w_i} & \text{if } s \geq 2 \\ E[y^j | y = z] f(z) = \frac{\sum_{i=1}^n w_i K(z - y_i) y_i^j}{\sum_{i=1}^n w_i} & \text{if } s = 1 \end{cases}$$

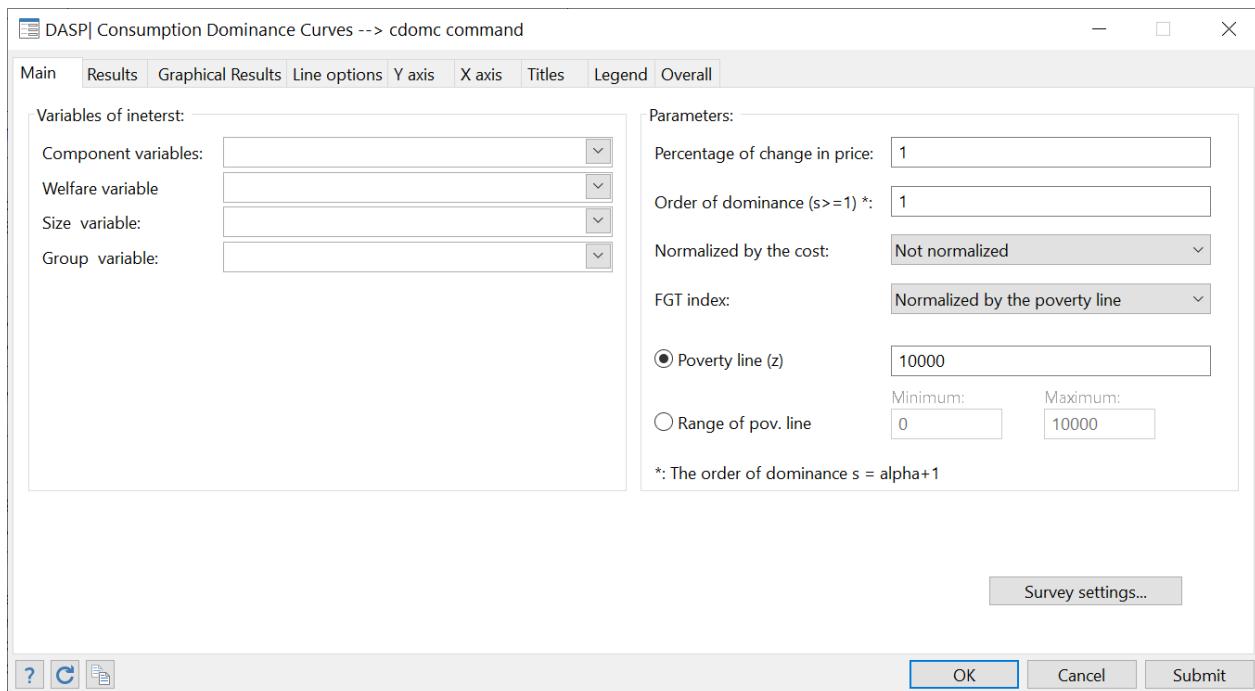
where  $K()$  is a kernel function and  $y^j$  is the  $j^{th}$  commodity. Dominance of order  $s$  is checked by setting  $\alpha = s - 1$ . The **cdomc** module draws such curves easily. The module can:

- draw more than one CD curve simultaneously whenever more than one component is selected.

- draw the CD curves with confidence intervals.
- estimate the impact of change in price of a given component on FGT index (CD curve) for a specified poverty line.
- draw the normalised CD curves by the average of the component.
- list or save the coordinates of the curves.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **cdomc**, type the command *db cdomc* in the command window.

**Figure 21: Consumption dominance curves dialog box**



## 16.7 Difference/Ratio between consumption dominance curves (cdomc2d)

The **cdomc2d** module draws difference or ratio between consumption dominance curves and their associated confidence intervals by taking sampling design into account. The module can:

- draw differences between consumption dominance curves and associated two-sided, lower-bounded or upper-bounded confidence intervals.
- list or save the coordinates of the differences and their confidence intervals.
- save the graphs in different formats:
  - \*.gph : Stata format.
  - \*.wmf : typically recommended to insert graphs in Word documents.

- \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

## 17 DASP and the progressivity

### 17.1 DASP and the progressivity indices

The **iprog** module estimates the progressivity indices for a given list of variables (components). Progressivity indices are:

- The Kakwani index; and
- The Reynolds and Smolensky index.

### 17.2 DASP and the progressivity curves

#### 17.2.1 Checking the progressivity of taxes or transfers

The module **cprog** allows checking whether taxes or transfers are progressive.

Let  $X$  be a gross income,  $T$  be a given tax and  $B$  be a given transfer.

The tax  $T$  is Tax Redistribution (TR) progressive if :

$$PR(p) = L_X(p) - C_T(p) > 0 \quad \forall p \in ]0,1[$$

The transfer  $B$  is Tax Redistribution (TR) progressive if :

$$PR(p) = C_B(p) - L_X(p) > 0 \quad \forall p \in ]0,1[$$

The tax  $T$  is Income Redistribution (IR) progressive if :

$$PR(p) = C_{X-T}(p) - L_X(p) > 0 \quad \forall p \in ]0,1[$$

The transfer  $B$  is Income Redistribution (IR) progressive if :

$$PR(p) = C_{X+B}(p) - L_X(p) > 0 \quad \forall p \in ]0,1[$$

#### 17.2.2 Checking the progressivity of transfer vs tax

The module **cprogbt** allows checking whether a given transfer is more progressive than a given tax.

The transfer  $B$  is more Tax Redistribution (TR) progressive than a tax  $T$  if :

$$PR(p) = C_B(p) + C_T(p) - 2L_X(p) > 0 \quad \forall p \in ]0,1[$$

The transfer  $B$  is more Income Redistribution (TR) progressive than a tax  $T$  if :

$$PR(p) = C_{X+B}(p) - C_{X-T}(p) > 0 \quad \forall p \in ]0,1[$$

## 18 Dominance

### 18.1 Poverty dominance (dompov)

Distribution 1 dominates distribution 2 at order s over the range  $[z^-, z^+]$  if only if:

$$P_1(\zeta; \alpha) < P_2(\zeta; \alpha) \quad \forall \quad \zeta \in [z^-, z^+] \text{ for } \alpha = s - 1.$$

This involves comparing stochastic dominance curves at order  $s$  or FGT curves with  $\alpha = s - 1$ . This application estimates the points at which there is a reversal of the ranking of the curves. Said differently, it provides the crossing points of the dominance curves, that is, the values of  $\zeta$  and  $P_1(\zeta; \alpha)$  for which  $P_1(\zeta; \alpha) = P_2(\zeta; \alpha)$  when:

$\text{sign}(P_1(\zeta - \eta; \alpha) - P_2(\zeta - \eta; \alpha)) = \text{sign}(P_2(\zeta + \eta; \alpha) - P_1(\zeta + \eta; \alpha))$  for a small  $\eta$ . The crossing points  $\zeta$  can also be referred to as "critical poverty lines".

The **dompov** module can be used to check for poverty dominance and to compute critical values. This module is mostly based on Araar (2006):

Araar, Abdelkrim, (2006), [Poverty, Inequality and Stochastic Dominance, Theory and Practice: Illustration with Burkina Faso Surveys](#), Working Paper: 06-34. CIRPEE, Department of Economics, Université Laval.

Interested users are encouraged to consider the exercises that appear in Section 23.9.

### 18.2 Inequality dominance (domineq)

Distribution 1 inequality-dominates distribution 2 at the second order if and only if:

$$L_1(p) \leq L_2(p) \quad \forall \quad p \in [0,1]$$

The module **domineq** can be used to check for such inequality dominance. It is based mainly on Araar (2006):

Araar, Abdelkrim, (2006), [Poverty, Inequality and Stochastic Dominance, Theory and Practice: Illustration with Burkina Faso Surveys](#), Working Paper: 06-34. CIRPEE, Department of Economics, Université Laval.

Intersections between curves can be estimated with this module. It can also used to check for tax and transfer progressivity by comparing Lorenz and concentration curves.

### 18.3 DASP and bi-dimensional poverty dominance (`dombdpov`)

Let two dimensions of well-being be denoted by  $k = 1, 2$ . The intersection bi-dimensional FGT index for distribution  $D$  is estimated as

$$\hat{P}_D(Z; A) = \frac{\sum_{i=1}^n w_i \left[ \prod_{k=1}^2 (z^k - y_i^k)_+^{\alpha_k} \right]}{\sum_{i=1}^n w_i}$$

where  $Z = (z_1, z_2)$  and  $A = (\alpha_1, \alpha_2)$  are vectors of poverty lines and parameters  $\alpha$  respectively, and  $x_+ = \max(x, 0)$ .

Distribution 1 dominates distribution 2 at orders  $(s_1, s_2)$  over the range  $[0, Z^+]$  if and only if:

$$P_1(Z; A = s - 1) < P_2(Z; A = s - 1) \quad \forall \quad Z \in [0, z_1^+] \times [0, z_2^+] \text{ and for } \alpha_1 = s_1 - 1, \quad \alpha_2 = s_2 - 1.$$

The DASP `dombdpov` module can be used to check for such dominance.

For each of the two distributions:

- The two variables of interest (dimensions) should be selected.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided, and the level of confidence used can be changed.
- Surfaces showing the difference, the lower bound and the upper bound of the confidence surfaces are plotted interactively with the GnuPlot tool.
- Coordinates can be listed.
- Coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Section [23.14](#).

## 19 Distributive tools

### 19.1 Quantile curves (c\_quantile)

The quantile at a percentile  $p$  of a continuous population is given by:

$Q(p) = F^{-1}(p)$  where  $p = F(y)$  is the cumulative distribution function at  $y$ .

For a discrete distribution, let  $n$  observations of living standards be ordered such that  $y_1 \leq y_2 \leq \dots \leq y_i \leq y_{i+1} \leq \dots \leq y_n$ . If  $F(y_i) < p \leq F(y_{i+1})$ , we define  $Q(p) = y_{i+1}$ . The normalised quantile is defined as  $\bar{Q}(p) = Q(p)/\mu$ .

Interested users are encouraged to consider the exercises that appear in Section 23.13.

### 19.2 Income share and cumulative income share by group quantiles (quinsh)

This module can be used to estimate the income shares, as well as, the cumulative income shares by quantile groups. The user can indicate the number of group partition. For instance, if the number is five, the quintile income shares are provided. We can also plot the graph bar of the estimated income shares.

### 19.3 Density curves (cdensity)

The Gaussian kernel estimator of a density function  $f(x)$  is defined by

$$\hat{f}(x) = \frac{\sum_i w_i K_i(x)}{\sum_{i=1}^n w_i} \quad \text{and} \quad K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp(-0.5 \lambda_i(x)^2) \quad \text{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

where  $h$  is a bandwidth that acts as a “smoothing” parameter.

Interested users are encouraged to consider the exercises that appear in Section 23.13.

#### Boundary bias correction:

A problem occurs with kernel estimation when a variable of interest is bounded. It may be for instance that consumption is bounded between two bounds, a minimum and a maximum, and that we wish to estimate its density “close” to these two bounds. If the true value of the density at these two bounds is positive, usual kernel estimation of the density close to these two bounds will be biased. A similar problem occurs with nonparametric regressions.

#### Renormalisation approach:

One way to alleviate these problems is to use a smooth “corrected” Kernel estimator, following a paper by Peter Bearse, Jose Canals and Paul Rilstone. A boundary-corrected Kernel density estimator can then be written as

$$\hat{f}(x) = \frac{\sum_i w_i K_i^*(x) K_i(x)}{\sum_{i=1}^n w_i}$$

where

$$K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp(-0.5 \lambda_i(x)^2) \quad \text{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

and where the scalar  $K_i^*(x)$  is defined as

$$\begin{aligned} K_i^*(x) &= \psi(x)' P(\lambda_i(x)) \\ P(\lambda) &= \begin{pmatrix} 1 & \lambda & \frac{\lambda^2}{2!} & \cdots & \frac{\lambda^{s-1}}{(s-1)!} \end{pmatrix} \\ \psi(x) &= M^{-1} l_s' = \left( \int_A^B K(\lambda) P(\lambda) P(\lambda)' d\lambda \right)^{-1} l_s : A = \frac{x - \min}{h}, \quad B = \frac{x - \max}{h}, \quad l_s' \\ &= (1 \quad 0 \quad 0 \cdots 0) \end{aligned}$$

$\min$  is the minimum bound, and  $\max$  is the maximum one.  $h$  is the usual bandwidth. This correction removes bias to order  $h^s$ .

DASP offers four options, without correction, and with correction of order 1, 2 and 3.

## References

- Jones, M. C. 1993, simply boundary correction for Kernel density estimation. *Statistics and Computing* 3: 135-146.
- Bearse, P., Canals, J. and Rilstone, P. Efficient Semi parametric Estimation of Duration Models With Unobserved Heterogeneity, *Econometric Theory*, 23, 2007, 281-308

## Reflection approach

The reflection estimator approaches the boundary estimator by “reflecting” the data at the boundaries:

$$\begin{aligned} \hat{f}(x) &= \frac{\sum_i w_i K_i^r(x)}{\sum_{i=1}^n w_i} \\ K^r(x) &= K\left(\frac{x - X}{h}\right) + K\left(\frac{x + X - 2\min}{h}\right) K\left(\frac{x + X - 2\max}{h}\right) \end{aligned}$$

## References

- Cwik and Mielniczuk (1993), *Data-dependent Bandwidth Choice for a Grade Density Kernel Estimate*. *Statistics and probability Letters* 16: 397-405

Silverman, B. W. (1986), *Density for Statistics and Data Analysis*. London Chapman and Hall (p 30).

## 19.4 Nonparametric regression curves (cnpe)

Nonparametric regression is useful to show the link between two variables without specifying beforehand a functional form. It can also be used to estimate the local derivative of the first variable with respect to the second without having to specify the functional form linking them. Regressions with the **cnpe** module can be performed with one of the following two approaches:

### 19.4.1 Nadaraya-Watson approach

A Gaussian kernel regression of  $y$  on  $x$  is given by:

$$E(y|x) = \Phi(y|x) = \frac{\sum_i w_i K_i(x) y_i}{\sum_i w_i K_i(x)}$$

From this, the derivative of  $\Phi(y|x)$  with respect to  $x$  is given by

$$E\left(\frac{dy}{dx}|x\right) = \frac{\partial \Phi(y|x)}{\partial x}$$

### 19.4.2 Local linear approach

The local linear approach is based on a local OLS estimation of the following functional form:

$$K_i(x)^{1/2} y_i = \mu(x) K_i(x)^{1/2} + \mu'(x) K_i(x)^{1/2} (x_i - x) + \nu$$

or, alternatively, of:

$$K_i(x)^{1/2} y_i = \alpha K_i(x)^{1/2} + \beta K_i(x)^{1/2} (x_i - x) + \nu_i$$

Estimates are then given by:

$$E(y|x) = \alpha, E\left(\frac{dy}{dx}|x\right) = \beta$$

Interested users are encouraged to consider the exercises that appear in Section 23.13.

## 19.5 DASP and joint density functions

The module **sjdensity** can be used to draw a joint density surface. The Gaussian kernel estimator of the joint density function  $f(x, y)$  is defined as:

$$\hat{f}(\bar{x}, \bar{y}) = \frac{1}{2\pi h_x h_y \sum_{i=1}^n w_i} \sum_{i=1}^n w_i \exp\left(-\left(\frac{1}{2}\right)\left(\left(\frac{\bar{x} - x_i}{h_x}\right)^2 + \left(\frac{\bar{y} - y_i}{h_y}\right)^2\right)\right)$$

With this module:

- The two variables of interest (dimensions) should be selected;
- specific population subgroup can be selected;
- surfaces showing the joint density function are plotted interactively with the GnuPlot tool;
- coordinates can be listed;
- coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Subsection 23,14.

## 19.6 DASP and joint distribution functions

The module **sjdistrub** can be used to draw joint distribution surfaces. The joint distribution function  $F(\bar{x}, \bar{y})$  is defined as:

$$\hat{F}(\bar{x}, \bar{y}) = \frac{\sum_{i=1}^n w_i I(x_i \leq \bar{x}) I(y_i \leq \bar{y})}{\sum_{i=1}^n w_i}$$

With this module:

- The two variables of interest (dimensions) should be selected;
- specific population subgroups can be selected;
- surfaces showing the joint distribution function are plotted interactively with the GnuPlot tool;
- coordinates can be listed;
- coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in 23,14.

## 20 DASP and pro-poor growth

### 20.1 DASP and pro-poor indices

The module **ipropoor** estimates simultaneously the three following pro-poor indices:

1. *The Chen and Ravallion pro-poor index (2003):*

$$\text{Index} = \frac{W_1(z) - W_2(z)}{F_1(z)}$$

where  $W_D(z)$  is the Watts index for distribution  $D \in [1,2]$  and  $F_1(z)$  is the headcount for index for the first distribution, both with poverty lines  $z$ .

2. *The Kakwani and Pernia pro-poor index (2000):*

$$\text{Index} = \frac{P_1(z, \alpha) - P_2(z\alpha)}{P_1(z, \alpha) - P_1(z(\mu_1/\mu_2), \alpha)}$$

3. *The Kakwani, Khandker and Son pro-poor index (2003):*

$$\text{Index\_1} = g \frac{P_1(z, \alpha) - P_2(z\alpha)}{P_1(z, \alpha) - P_1(z(\mu_1/\mu_2), \alpha)}$$

where the average growth is  $g(\mu_2 - \mu_1)/\mu_1$  and where a second index is given by:

$$\text{Index\_2} = \text{Index\_1} - g$$

- One variable of interest should be selected for each distribution.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.
- A level for the parameter  $\alpha$  can be chosen for each of the two distributions.

## 20.2 DASP and pro-poor curves

Pro-poor curves can be drawn using either the primal or the dual approach. The former uses income levels. The latter is based on percentiles.

### 20.2.1 Growth incidence curve (gicur)

Let:

$Q_t(p)$  : quantile at percentile  $p$  at period  $t$ . The GIC( $p$ ) curve is:

$$GIC(p) = \frac{Q_2(p) - Q_1(p)}{Q_1(p)}$$

### 20.2.2 Primal pro-poor curves

The change in the distribution from state 1 to state 2 is s-order absolutely pro-poor with standard **cons** if:

$$\Delta(z, s) = (P_2(z + cons, \alpha = s - 1) - P_1(z, \alpha = s - 1)) < 0 \quad \forall z \in [0, z^+]$$

The change in the distribution from state 1 to state 2 is s-order relatively pro-poor if:

$$\Delta(z, s) = z \left( P_2(z \frac{\mu_2}{\mu_1}, \alpha = s - 1) - P_1(z, \alpha = s - 1) \right) < 0 \quad \forall z \in [0, z^+]$$

The module **cpropoorp** can be used to draw these primal pro-poor curves and their associated confidence interval by taking into account sampling design. The module can:

- draw pro-poor curves and their two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences between the curves as well as those of the confidence intervals;
- save the graphs in different formats:
  - \*.gph : Stata format;
  - \*.wmf : typically recommended to insert graphs in Word documents;
  - \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section 23.15.

### 20.2.3 Dual pro-poor curves

Let:

$Q(p)$  : quantile at percentile  $p$ .

$GL(p)$  : Generalised Lorenz curve at percentile  $p$ .

$\mu$  : average living standards.

The change in the distribution from state 1 to state 2 is first order absolutely pro-poor with standard  $cons=o$  if:

$$\Delta(z, s) = Q_2(p) - Q_1(p) > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

or equivalently if:

$$\Delta(z, s) = \frac{Q_2(p) - Q_1(p)}{Q_1(p)} > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

The change in the distribution from state 1 to state 2 is first order relatively pro-poor if:

$$\Delta(z, s) = \frac{Q_2(p)}{Q_1(p)} - \frac{\mu_2}{\mu_1} > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

The change in the distribution from state 1 to state 2 is second order absolutely pro-poor if:

$$\Delta(z, s) = GL_2(p) - GL_1(p) > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

or equivalently if:

$$\Delta(z, s) = \frac{GL_2(p) - GL_1(p)}{GL_1(p)} > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

The change in the distribution from state 1 to state 2 is first order relatively pro-poor if:

$$\Delta(z, s) = \frac{GL_2(p)}{GL_1(p)} - \frac{\mu_2}{\mu_1} > 0 \quad \forall p \in [0, \dots, p^+] = F(z^+)]$$

The module **cpropoord** can be used to draw these dual pro-poor curves and their associated confidence interval by taking into account sampling design. The module can:

- draw pro-poor curves and their two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences between the curves as well as those of the confidence intervals;
- save the graphs in different formats:

- \*.gph : Stata format;
- \*.wmf : typically recommended to insert graphs in Word documents;
- \*.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section 23.15

## 21 DASP and Benefit Incidence Analysis

### 21.1 Benefit incidence analysis

The main objective of benefit incidence is to analyse the distribution of benefits from the use of public services according to the distribution of living standards.

Two main sources of information are used. The first informs on access of household members to public services. This information can be found in usual household surveys. The second deals with the amount of total public expenditures on each public service. This information is usually available at the national level and sometimes in a more disaggregated format, such as at the regional level. The benefit incidence approach combines the use of these two sources of information to analyse the distribution of public benefits and its progressivity.

Formally, let

- $w_i$  be the sampling weight of observation  $i$ ;
- $y_i$  be the living standard of members belonging to observation  $i$  (i.e., *per capita* income);
- $e_i^s$  be the number of “eligible” members of observation  $i$ , i.e., members that “need” the public service provided by sector  $s$ . There are  $S$  sectors;
- $f_i^s$  be the number of members of observation  $i$  that effectively use the public service provided by sector  $s$ ;
- $g_i$  be the socio-economic group of eligible members of observation  $i$  (typically classified by income percentiles);
- $c_i$  be a subgroup indicator for observation  $i$  (e.g., 1 for a rural resident, and 2 for an urban resident). Eligible members can thus be grouped into population exclusive subgroups;

$E_r^s$  be total public expenditures on sector  $s$  in area  $r$ . There are  $R$  areas (the area here refers to the geographical division which one can have reliable information on total public expenditures on the studied public service);

$E^s$  be total public expenditures on sectors ( $E^s = \sum_{r=1}^R E_r^s$ ).

Here are some of the statistics that can be computed.

1. The share of a group  $g$  in sector  $s$  is defined as follows:

$$SH_g^s = \frac{\sum_{i=1}^n w_i f_i^s I(i \in g)}{\sum_{i=1}^n w_i f_i^s}$$

Note that:  $\sum_{g=1}^G SH_g^s = 1$ .

2. The rate of participation of a group  $g$  in sector  $s$  is defined as follows:

$$CR_g^s = \frac{\sum_{i=1}^n w_i f_i^s I(i \in g)}{\sum_{i=1}^n w_i e_i^s I(i \in g)}$$

This rate cannot exceed 100% since  $f_i^s \leq e_i^s \forall i$ .

3. The unit cost of a benefit in sector  $s$  for observation  $j$ , which refers to the household members that live in area  $r$ :

$$UC_j^s = \frac{E_r^s}{\sum_{j=1}^{n_r} w_j f_j^s}$$

where  $n_r$  is the number of sampled households in area  $r$ .

4. The benefit of observation  $i$  from the use of public sector  $s$  is:

$$B_i^s = f_i^s UC_i^s$$

5. The benefit of observation  $i$  from the use of the  $S$  public sectors is:

$$B_i = \sum_{s=1}^S B_i^s$$

6. The average benefit at the level of those eligible to a service from sector  $s$  and for those observations that belong to a group  $g$ , is defined as:

$$ABE_g^s = \frac{\sum_{i=1}^n w_i B_i^s I(i \in g)}{\sum_{i=1}^n w_i e_i^s I(i \in g)}$$

7. The average benefit for those that use the service  $s$  and belong to a group  $g$  is defined as:

$$ABF_g^s = \frac{\sum_{i=1}^n w_i B_i^s I(i \in g)}{\sum_{i=1}^n w_i f_i^s I(i \in g)}$$

8. The proportion of benefits from the service from sector  $s$  that accrues to observations that belong to a group  $g$  is defined as:

$$PB_g^s = \frac{B_g^s}{E^s}$$

where  $B_g^s = \sum_{i=1}^n w_i B_i^s I(i \in g)$ .

These statistics can be restricted to specific socio-demographic groups (e.g., rural/urban) by replacing  $I(i \in g)$  by  $I(i \in c)$ .

The **bian.ado** module allows the computation of these different statistics.

***Some characteristics of the module:***

- Possibility of selecting between one and six sectors.
- Possibility of using frequency data approach when information about the level of total public expenditures is not available.
- Generation of benefit variables by the type of public services (ex: primary, secondary and tertiary education levels) and by sector.
- Generation of unit cost variables for each sector.
- Possibility of computing statistics according to groups of observations.
- Generation of statistics according to social-demographic groups, such as quartiles, quintiles or deciles.

Public expenditures on a given service often vary from one geographical or administrative area to another. When information about public expenditures is available at the level of areas, this information can be used with the **bian** module to estimate unit cost more accurately.

**Example 1**

Observation i	HH size	Eligible HH members	Frequency	Area indicator	Total level of regional public expenditures
1	7	3	2	1	14000
2	4	2	2	1	14000
3	5	5	3	1	14000
4	6	3	2	2	12000
5	4	2	1	2	12000

In this example, the first observation contains information on household 1.

- This household contains 7 individuals;
- Three individuals in this household are eligible to the public service;
- Only 2 among the 3 eligible individuals benefit from the public service;
- This household lives in area 1. In this area, the government spends a total of 14000 to provide the public service for the 7 users of this area (2+2+3).

The unit cost in area 1 equals:  $14000/7=2000$

The unit cost in area 2 equals:  $12000/3=4000$

By default, the area indicator is set to 1 for all households. When this default is used, the variable Regional public expenditures (the fifth column that appears in the dialog box) should be set to total public expenditures at the national level. This would occur when the information on public expenditures is only available at the national level.

### Example 2

Observation i	HH size	Eligible members	Frequency	Area indicator	Regional public expenditures
1	7	3	2	1	28000
2	4	2	2	1	28000
3	5	5	3	1	28000
4	6	3	2	1	28000
5	4	2	1	1	28000

The unit cost benefit (at the national level) equals:  $28000/10=2800$

Interested users are encouraged to consider the exercises that appear in Section [23.16](#)

## 21.2 Marginal benefit incidence analysis (imbi)

Despite the simplicity benefit incidence, many researchers has addressed a serious criticisms to this approach when results are used to project the change in benefit incidence with an expansion in public spending. Indeed, BA describes the actual distribution of benefits and not the expected change in participation with an expansion in a given public service. Studies of marginal benefit incidence are devoted to respond to such problematic.

With cross-section data, to estimate the marginal benefit incidence, that is, who gains from an expansion in the program or service, it is useful to exploit the geographic variation in access both between households and between regions as a source of information for understanding the diffusion process that generates access.

First, assume that a given country with  $i = 1, \dots, N$  regions. Assume also that households are ranked by welfare indicator and assigned to one of  $q = 1, \dots, Q$  income intervals (Quintile for instance). Then, the ranking is done at regional level. Let  $x_{ij}^q$  be the benefit incidence –or frequency- of a program or service in household  $j$  belonging to interval  $q$  and living in region  $i$ . The average benefit incidence in interval  $q$  for households in region  $i$  is denoted by  $X_i^q$ , and the overall region average is  $i$  denoted by  $X_i$ . If  $J_i^q$  is the number of households in interval  $q$  for region  $i$ , the two averages are respectively equal to the following:

$$X_i^q = \sum_{j=1}^{J_i^q} x_{ij}^q / J_i^q \quad (\text{A.1})$$

$$X_i = \sum_{q=1}^Q \sum_{j=1}^{J_i^q} x_{ij}^q / \sum_{q=1}^Q J_i^q \quad (\text{A.2})$$

### 21.2.1 Derivative of the linear locally estimator approach

We use the LLE non-parametric approach, which is defined in section 18.4.2, we simply estimate the derivative  $\frac{\partial X_i}{\partial X_i^q} \Big|_{X_i}$ .

### 21.2.2 Ajwad and Quentin (2001) approach

The authors propose to regress the incidence in each of the intervals in the regions against the regional means, using  $Q$  regressions:

$$X_i^q = \alpha^q + \beta^q \left( \frac{\sum_{q=1, j=1}^{Q, J_i^q} x_{ij}^q - \sum_{j=1}^{J_i^q} x_{ij}^q}{\sum_{q=1}^Q J_i^q - J_i^q} \right) + \varepsilon^q \quad \text{for } q = 1, \dots, Q \quad (\text{A.3})$$

To avoid endogeneity means at regional level are for all households, except for those belonging to interval  $q$ . Pooling all observations from the various intervals together, they estimate one regression:

$$X_i^q = \sum_{q=1}^Q \alpha^q + \sum_{q=1}^Q \beta^q \left( \frac{\sum_{q=1, j=1}^{Q, J_i^q} x_{ij}^q - \sum_{j=1}^{J_i^q} x_{ij}^q}{\sum_{q=1}^Q J_i^q - J_i^q} \right) + \varepsilon^q \quad (\text{A.4})$$

Since the average marginal increase in access from a unitary increase in mean access is one, this implies the following restriction:

$$\sum_{q=1}^Q \frac{\beta^q}{Q - 1 + \beta^q} = 1 \quad (\text{A.5})$$

Writing  $\beta^q$ , the parameter for interval  $Q$  in relation to the other parameters, yields the following:

$$\beta^q = \frac{(Q - 1) \left( 1 - \sum_{q=1}^Q \frac{\beta^q}{Q - 1 + \beta^q} \right)}{\sum_{q=1}^Q \frac{\beta^q}{Q - 1 + \beta^q}} \quad (\text{A.6})$$

To take into account the restriction (A.6), (A.4) is estimated with nonlinear least squares. It can also be shown that a change in benefit incidence for the households belonging to quintile  $q$  in response to an increase in the aggregate incidence is as follows:

$$\frac{\partial X_i}{\partial X_i^q} = \frac{Q \beta^q}{Q - 1 + \beta^q} \quad (\text{A.7})$$

The module imbi.ado allows estimating the marginal benefit incidence with one of the presented two approaches above. The user will indicate:

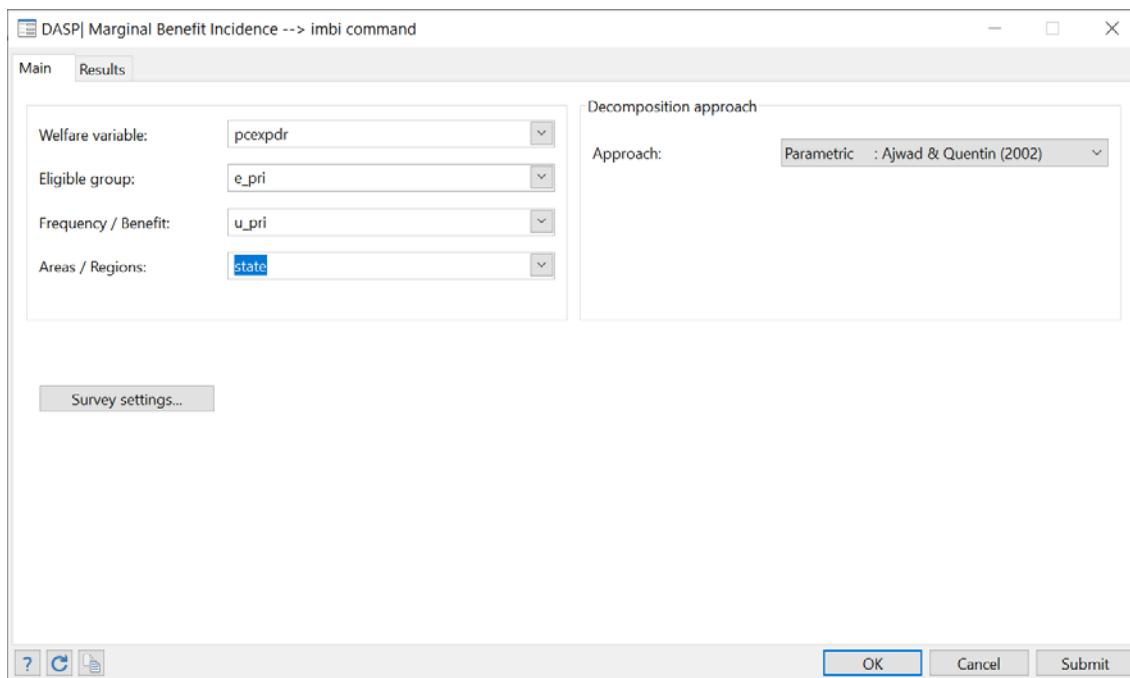
- The welfare indicator, like the per capita expenditures:

- The variable eligible members. For instance, a household with two children aged between 6 and 11 years will have 2 members eligible to the primary school.
- The variable Frequency or Benefit: here the user can indicate the number of household members that effectively use the service, or the estimated monetary benefit from the public service.
- The variable region: this variable indicates the region (integer value) in which the household lives. It is recommended to select the appropriate level of spatial unit to ensure the representativeness of households at the regional level.

### Example

Using the Nigerian household survey of 2004, we estimate the MBI for the public primary school service based on Ajwad and Quentin (2001) approach.

**Figure 22: IMBI dialog box**



```
. imbi u_pri e_pri, welfare(pcexpdr) hregion(state) appr(aq) drr(0)
```

---

Sampling weight : popwt  
 Regional variable : state  
 Number of regions : 37

---

Marginal Benefit Incidence  
 Estimated with the Ajwad & Quentin approach

Group	alpha	beta	marginBen
Quantile_1	-0.284582	1.203984	1.156791
Quantile_2	-0.057060	1.032782	1.026055
Quantile_3	-0.023669	1.038546	1.030601
Quantile_4	0.091684	0.955386	0.963988
Quantile_5	0.255963	0.787628	0.822566

## 22 Disaggregating grouped data

The **ungroup** DASP module generates disaggregated data from aggregate distributive information. Aggregate information is obtained from cumulative income shares (or Lorenz curve ordinates) at some percentiles. For instance:

Percentile (p)	0.10	0.30	0.50	0.60	0.90	1.00
Lorenz values: L(p)	0.02	0.10	0.13	0.30	0.70	1.00

The user must specify the total number of observations to be generated. The user can also indicate the number of observations to be generated specifically at the top and/or at the bottom of the distribution, in which case the proportion (in %) of the population found at the top or at the bottom must also be specified.

Remarks:

- If only the total number of observations is set, the generated data are self-weighted (or uniformly distributed over percentiles).
- If a number of observations is set for the bottom and/or top tails, the generated data are not self-weighted and a weight variable is provided in addition to the generated income variable.
  - Example: Assume that the total number of observations to be generated is set to 1900, but that we would like the bottom 10% of the population to be represented by 1000 observations. In this case, weights will equal 1/1000 for the bottom 1000 observations and 1/100 for the remaining observations (the sum of weights being normalised to one).
- The generated income vector takes the name of \_y and the vector weight, \_w.

- The number of observations to be generated does not have to equal the number of observations of the sample that was originally used to generate the aggregated data. The ungroup module cannot in itself serve to estimate the sampling errors that would have occurred had the original sample data been used to estimate poverty and/or inequality estimates.
- The user can select any sample size that exceeds (number\_of\_classes+1), but it may be more appropriate for statistical bias-reduction purposes to select relatively large sizes.

## **STAGE I Generating an initial distribution of incomes and percentiles**

### **S.1.1: Generating a vector of percentiles**

Starting from information on the importance of bottom and top groups and on the number of observations to be generated, we first generate a vector of percentiles.

#### **Examples**

##### **Notations**

NOBS: number of total observations

F: vector of percentiles

B\_NOBS: number of observations for the bottom group

T\_NOBS: number of observations for the top group.

- For NOBS=1000 spread equally across all percentiles,  $F=0.001, 0.002\dots 0.999, 1$ . To avoid the value  $F=1$  for the last generated observation, we can simply replace F by  $F-(0.5/NOBS)$ .
- For NOBS=2800, B\_NOBS=1000 and T\_NOBS=1000, with the bottom and top groups being the first and last deciles:
  - a.  $F=0.0001, 0.0002, \dots, 0.0999, 0.1000$  in 0001/1000
  - b.  $F=0.1010, 0.1020, \dots, 0.8990, 0.9000$  in 1001/1800
  - c.  $F=0.9001, 0.9002, \dots, 0.9999, 1.0000$  in 1801/2800

Adjustments can also be made to avoid the case of  $F(1)=1$ .

The weight vector can easily be generated.

### **S.1.2: Generating an initial distribution of incomes**

The user must indicate the form of distribution of the disaggregated data.

#### ***Normal and log normal distributions:***

Assume that  $x$  follows a lognormal distribution with mean  $\mu$  and variance  $\sigma^2$ . The Lorenz curve is defined as follows:

$$L(p) = \Phi\left(\frac{\ln(x) - (\mu - \sigma^2)}{\sigma}\right) \text{ and } p = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

We assume that  $\mu = 1$  and we estimate the variance using the procedure suggested by Shorrocks and Wan (2008): a value for the standard deviation of log incomes,  $\sigma$ , is obtained by averaging the  $m - 1$  estimates of  $\sigma^k = \Phi^{-1}(p_k) - \Phi^{-1}(L(p_k))$   $k = 1, \dots, m - 1$  where  $m$  is the number of classes and  $\Phi$  is the standard normal distribution function (Aitchison and Brown 1957; Kolenikov and Shorrocks 2005, Appendix).

### **Generalised Quadratic Lorenz Curve:**

It is assumed that:

$$L(1 - L) = a(p^2 - L) + bL(p - 1) + c(p - L)$$

We can regress  $L(1 - L)$  on  $(p^2 - L)$ ,  $L(p - 1)$  and  $(p - L)$  without an intercept, dropping the last observation since the chosen functional form forces the curve to go through  $(1,1)$ .

We have  $Q(p) = -\frac{b}{2} - \frac{(2mp+n)(mp^2+np+e^2)^{-0.5}}{4}$

$$\begin{aligned} e &= a + b + c + 1 \\ m &= b^2 - 4a \\ n &= -2be - 4c \end{aligned}$$

### **Beta Lorenz Curve:**

It is assumed that:

$$\log(p - L) = \log(\theta) + \gamma \log(p) + \delta \log(1 - p)$$

After estimating the parameters, we can generate quantiles as follows

$$Q(p) = \theta + p^\gamma (1 - p)^\delta \left[ \frac{\gamma}{p} - \frac{\delta}{(1 - p)} \right]$$

See also Datt (1998).

### **The Singh-Maddala distribution**

The distribution function proposed by Singh and Maddala (1976) takes the following form:

$$F(x) = 1 - \left[ \frac{1}{1 + (x/b)^a} \right]^q$$

where  $a \geq 0$ ,  $b \geq 0$ ,  $q \geq 1/a$  are parameters to be estimated. The income ( $x$ ) is assumed to be equal to or greater than zero. The density function is defined as follows:

$$f(x) = (aq/b)(1 + (x/b)^a)^{-(q+1)}(x/b)^{(a-1)}$$

Quantiles are defined as follows:

$$Q(p) = b((1 - p)^{-1/q} - 1)^{1/a}$$

We follow Jenkins (2008)'s approach for the estimation of parameters. For this, we maximize the likelihood function, which is simply the product of density functions evaluated at the average income of each class:

<http://stata-press.com/journals/stbcontents/stb48.pdf>

## **STAGE II Adjusting the initial distribution to match the aggregated data (optional).**

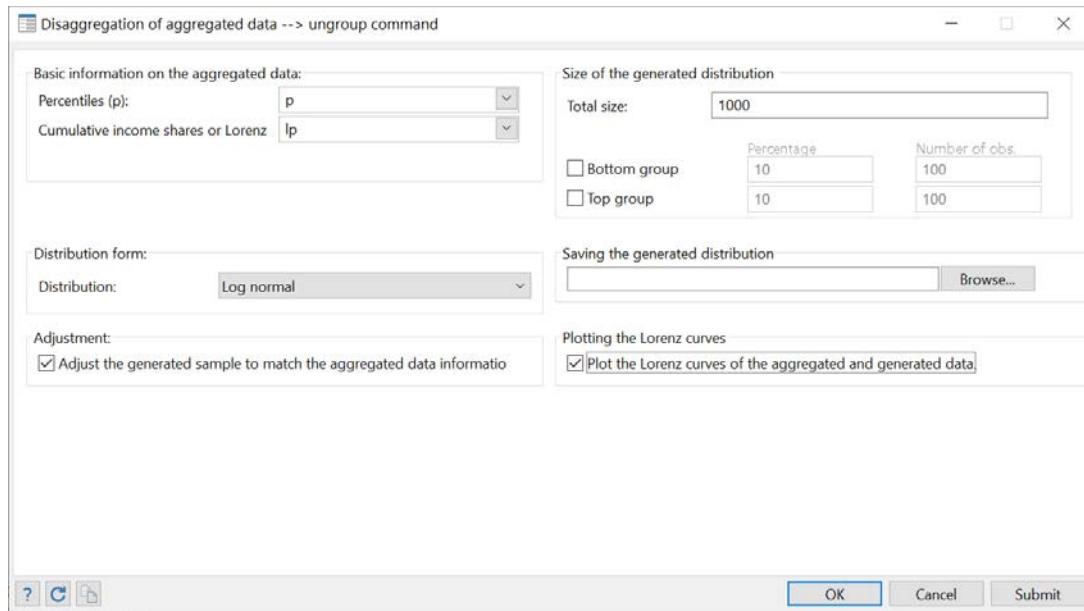
This stage adjusts the initial vector of incomes using the Shorrocks and Wan (2008) procedure. This procedure proceeds with two successive adjustments:

- **Adjustment 1:** Correcting the initial income vector to ensure that each income group has its original mean income.
- **Adjustment 2:** Smoothing the inter-class distributions.

The generated sample is saved automatically in a new Stata data file (called by default *ungroup\_data.dta*; names and directories can be changed). The user can also plot the Lorenz curves of the aggregated (when we assume that each individual has the average income of his group) and generated data.

Dialog box of the **ungroup** module

**Figure 23: ungroup dialog box**



## **Illustration with Burkina Faso household survey data**

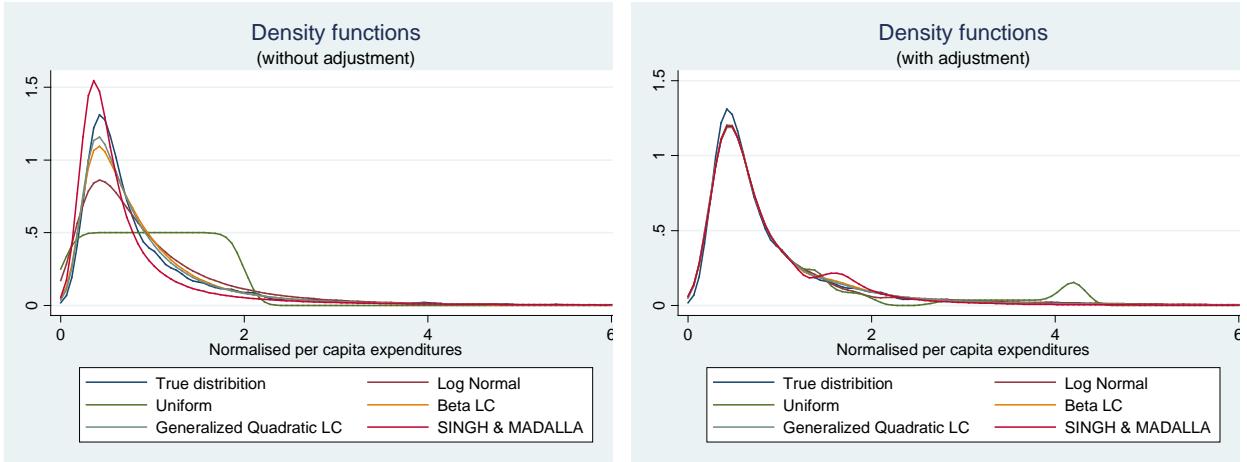
In this example, we use disaggregated data to generate aggregated information. Then, we compare the density curve of the true data with those of the data generated through disaggregation of the previously aggregated data.

```
gen fw=size*weight  
gen y=exppc/r(mean)  
clorenz y, hs(size) lres(1)
```

### **Aggregated information:**

p	L(p)
.1	.0233349
.2	.0576717

.3	.0991386
.4	.1480407
.5	.2051758
.6	.2729623
.7	.3565971
.8	.4657389
.9	.6213571
1	1.00000



## 23 Appendices

### 23.1 Appendix A: illustrative household surveys

#### 23.1.1 The 1994 Burkina Faso survey of household expenditures (bkf94I.dta)

This is a nationally representative survey, with sample selection using two-stage stratified random sampling. Seven strata were formed. Five of these strata were rural and two were urban. Primary sampling units were sampled from a list drawn from the 1985 census. The last sampling units were households.

List of variables

<i>strata</i>	Stratum in which a household lives
<i>psu</i>	Primary sampling unit
<i>weight</i>	Sampling weight
<i>size</i>	Household size
<i>exp</i>	Total household expenditures
<i>expeq</i>	Total household expenditures per adult equivalent
<i>expcp</i>	Total household expenditures per capita
<i>gse</i>	Socio-economic group of the household head 1 wage-earner (public sector) 2 wage-earner (private sector) 3 Artisan or trader 4 Other type of earner 5 Crop farmer 6 Subsistence farmer 7 Inactive
<i>sex</i>	Sex of household head 1 Male 2 Female
<i>zone</i>	Residential area 1 Rural 2 Urban

#### 23.1.2 The 1998 Burkina Faso survey of household expenditures (bkf98I.dta)

This survey is similar to the 1994 one, although ten strata were used instead of seven for 1994. To express 1998 data in 1994 prices, two alternative procedures have been used. First, 1998 expenditure data were

multiplied by the ratio of the 1994 official poverty line to the 1998 official poverty line:  $z_{1994}/z_{1998}$ . Second, 1998 expenditure data were multiplied by the ratio of the 1994 consumer price index to the 1998 consumer price index:  $ipc_{1994}/ipc_{1998}$ .

#### List of new variables

*expcpz* Total household expenditures per capita deflated by ( $z_{1994}/z_{1998}$ )

*expcpi* Total expenditures per capita deflated by ( $ipc_{1994}/ipc_{1998}$ )

### 23.1.3 Canadian Survey of Consumer Finance (a subsample of 1000 observations – can6.dta)

#### List of variables

*X* Yearly gross income per adult equivalent.

*T* Income taxes per adult equivalent.

*B<sub>1</sub>* Transfer 1 per adult equivalent.

*B<sub>2</sub>* Transfer 2 per adult equivalent.

*B<sub>3</sub>* Transfer 3 per adult equivalent.

*B* Sum of transfers *B<sub>1</sub>*, *B<sub>2</sub>* and *B<sub>3</sub>*

*N* Yearly net income per adult equivalent (*X* minus *T* plus *B*)

### 23.1.4 Peru LSMS survey 1994 (A sample of 3623 household observations - PEREDE94I.dta)

#### List of variables

*expcc* Total expenditures, per capita (constant June 1994 soles per year).

*weight* Sampling weight

*size* Household size

*npubprim* Number of household members in public primary school

*npubsec* Number of household members in public secondary school

*npubuniv* Number of household members in public post-secondary school

### 23.1.5 Peru LSMS survey 1994 (A sample of 3623 household observations – PERU\_A\_I.dta)

List of variables

<i>hhid</i>	Household Id.
<i>exppc</i>	Total expenditures, per capita (constant June 1994 soles per year).
<i>size</i>	Household size
<i>literate</i>	Number of literate household members
<i>pliterate</i>	literate/size

### 23.1.6 The 1995 Colombia DHS survey (columbial.dta)

This sample is a part of the [Data from the Demographic and Health Surveys](#) (Colombia\_1995) which contains the following information for children aged 0-59 months

List of variables

<i>hid</i>	Household id
<i>haz</i>	height-for-age
<i>waz</i>	weight-for-age
<i>whz</i>	weight-for-height
<i>sprob</i>	survival probability
<i>wght</i>	sampling weight
<i>Asset</i>	asset index

### 23.1.7 The 1996 Dominican Republic DHS survey (Dominican\_republic1996I.dta)

This sample is a part of the [Data from the Demographic and Health Surveys](#) (Republic Dominican\_1996) which contains the following information for children aged 0-59 months

List of variables

<i>hid</i>	Household id
<i>haz</i>	height-for-age
<i>waz</i>	weight-for-age
<i>whz</i>	weight-for-height
<i>sprob</i>	survival probability
<i>wght</i>	sampling weight
<i>Asset</i>	asset index

## 23.2 Appendix B: labelling variables and values

- The following .do file can be used to set labels for the variables in **bkf94.dta**.
- For more details on the use of *label* command, type *help label* in the command window.

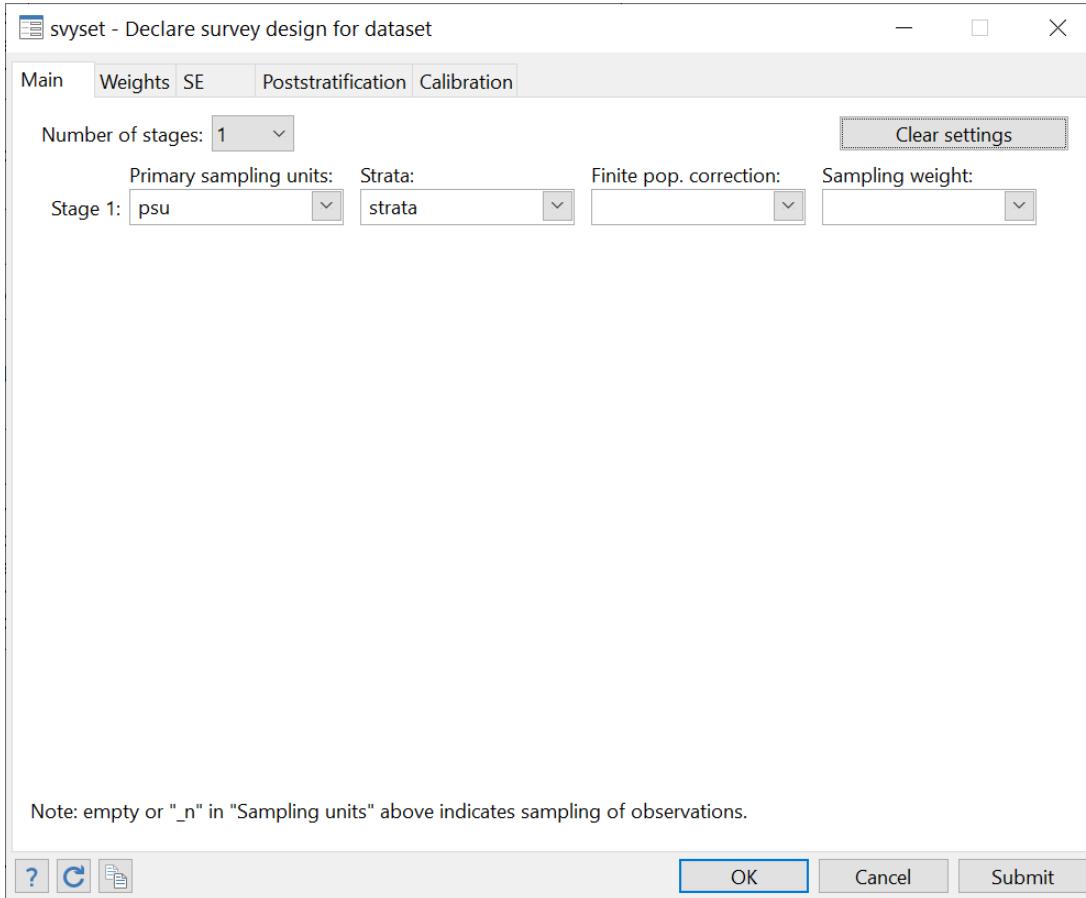
```
=====lab_bkf94.do =====
```

```
# delim ;  
  
/* To drop all label values */  
  
label drop _all;  
  
/* To assign labels */  
label var strata "Stratum in which a household lives";  
  
label var psu "Primary sampling unit";  
label var weight "Sampling weight";  
label var size "Household size";  
label var totexp "Total household expenditures";  
label var exppc "Total household expenditures per capita";  
label var expeq "Total household expenditures per adult equivalent";  
label var gse "Socio-economic group of the household head";  
  
/* To define the label values that will be assigned to the categorical variable gse */  
label define lvgse  
  
1 "wage-earner (public sector)"  
2 "wage-earner (private sector)"  
3 "Artisan or trader"  
4 "Other type of earner"  
5 "Crop farmer"  
6 "Subsistence farmer"  
7 "Inactive"  
;  
  
/*To assign the label values "lvgse" to the variable gse */  
label val gse lvgse;  
  
label var sex "Sex of household head";  
label def lvsex  
1 Male  
2 Female  
;  
  
label val sex lvsex;  
label var zone "Residential area";  
label def lvzone  
1 Rural  
2 Urban  
;  
  
label val zone lvzone;  
  
=====End=====
```

### 23.3 Appendix C: setting the sampling design

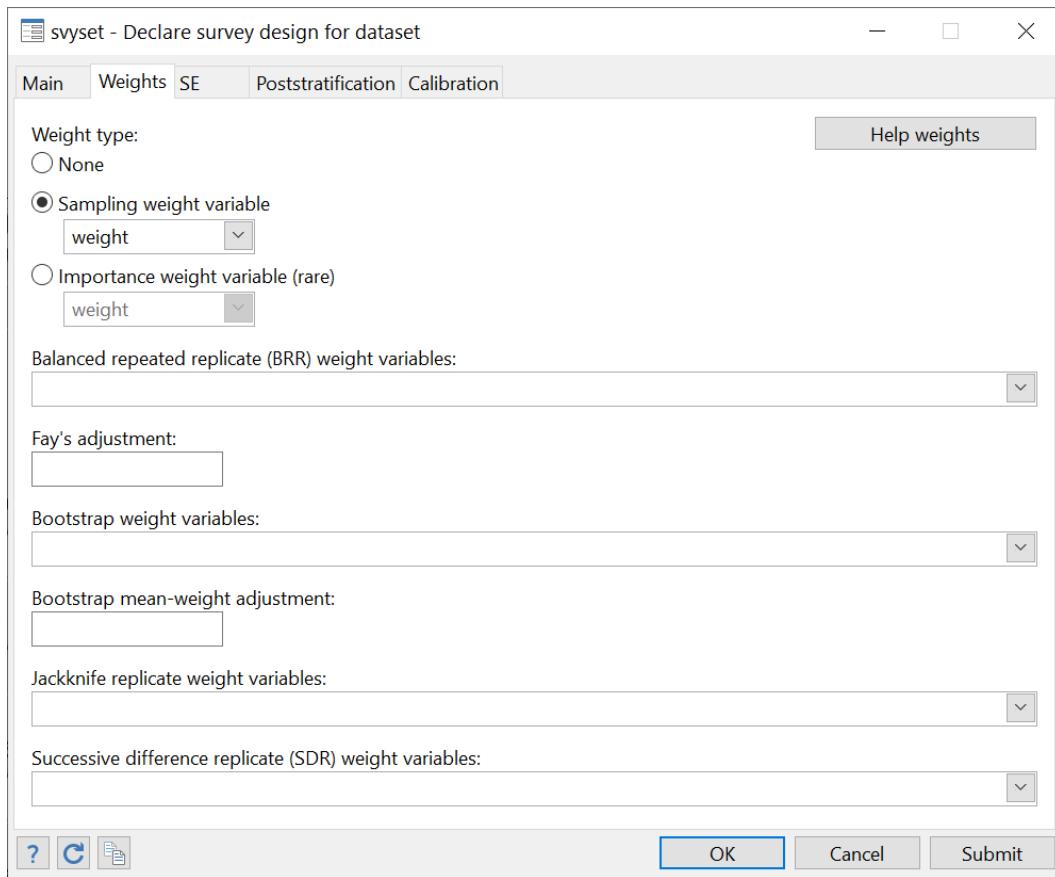
To set the sampling design for the data file **bkf94.dta**, open the dialog box for the command **svyset** by typing the syntax **db svyset** in the command window. In the Main panel, set STRATA and SAMPLING UNITS as follows:

Figure 24: Survey data settings



In the Weights panel, set SAMPLING WEIGHT VARIABLE as follows:

**Figure 25: Setting sampling weights**



Click on OK and save the data file.

To check if the sampling design has been well set, type the command `svydes`. The following will be displayed:

```
. svydes
Survey: Describing stage 1 sampling units

pweight: weight
          VCE: linearized
Single unit: missing
Strata 1: strata
          SU 1: psu
          FPC 1: <zero>

#Obs per Unit

```

Stratum	#Units	#Obs	min	mean	max
1	63	1,255	17	19.9	21
2	44	879	19	20.0	21
3	30	599	19	20.0	20
4	31	620	20	20.0	20
5	26	518	18	19.9	20
6	31	620	20	20.0	20
7	32	638	19	19.9	20
8	104	2,074	15	19.9	21
9	33	654	19	19.8	20
10	31	621	19	20.0	21
<b>10</b>	<b>425</b>	<b>8,478</b>	<b>15</b>	<b>19.9</b>	<b>21</b>

## Examples and exercises

### 23.4 Estimation of FGT poverty indices

"How poor was Burkina Faso in 1994?"

1. Open the **bkf94.dta** file and label variables and values using the information of Section 23.1.1.  
Type the *describe* command and then *label list* to list labels.
2. Use the information of Section 23.1.1. to set the sampling design and then save the file.
3. Estimate the headcount index using variables of interest *expcc* and *expeq*.
  - a. You should set *SIZE* to household size in order to estimate poverty over the population of individuals.
  - b. Use the so-called 1994 official poverty line of 41099 Francs CFA per year.
4. Estimate the headcount index using the same procedure as above except that the poverty line is now set to 60% of the median.
5. Using the official poverty line, how does the headcount index for male- and female-headed households compare?
6. Can you draw a 99% confidence interval around the previous comparison? Also, set the number of decimals to 4.

### Answers

#### Q.1

If **bkf94.dta** is saved in the directory **c:/data**, type the following command to open it:  
*use "C:\data\bkf94.dta", clear*

If **lab\_bkf94.do** is saved in the directory **c:/do\_files**, type the following command to label variables and labels:

*do "C:\do\_files\lab\_bkf94.do"*

Typing the command *describe*, we obtain:

obs:	8,625	vars:	9	31 Oct 2006 13:48
size:	285,087 (99.6% of memory free)			
storage display	value	variable	label	
variable	name	type	format	label
weight		float	%9.0g	Sampling weight
size		byte	%8.0g	Household size
strata		byte	%8.0g	Stratum in which a household lives
psu		byte	%8.0g	Primary sampling unit
gse		byte	%29.0g	gse
sex		byte	%8.0g	sex
zone		byte	%8.0g	zone
exp		double	%10.0g	Total household expenditures
expeq		double	%10.0g	Total household expenditures per adult equivalent
expcc		float	%9.0g	Total household expenditures per capita

Typing *label list*, we find:

zone:

- 1 Rural
- 2 Urban

sex:

- 1 Male
- 2 Female

gse:

- 1 wage-earner (public sector)
- 2 wage-earner (private sector)
- 3 Artisan or trader
- 4 Other type of earner
- 5 Crop farmer
- 6 Food farmer
- 7 Inactive

## Q.2

You can set the sampling design with a dialog box, as indicated in Section 23.3, or simply by typing

*svyset psu [pweight=weight], strata(strata) vce(linearized) singleunit(missing)*

Typing *svydes*, we obtain

Survey: Describing stage 1 sampling units

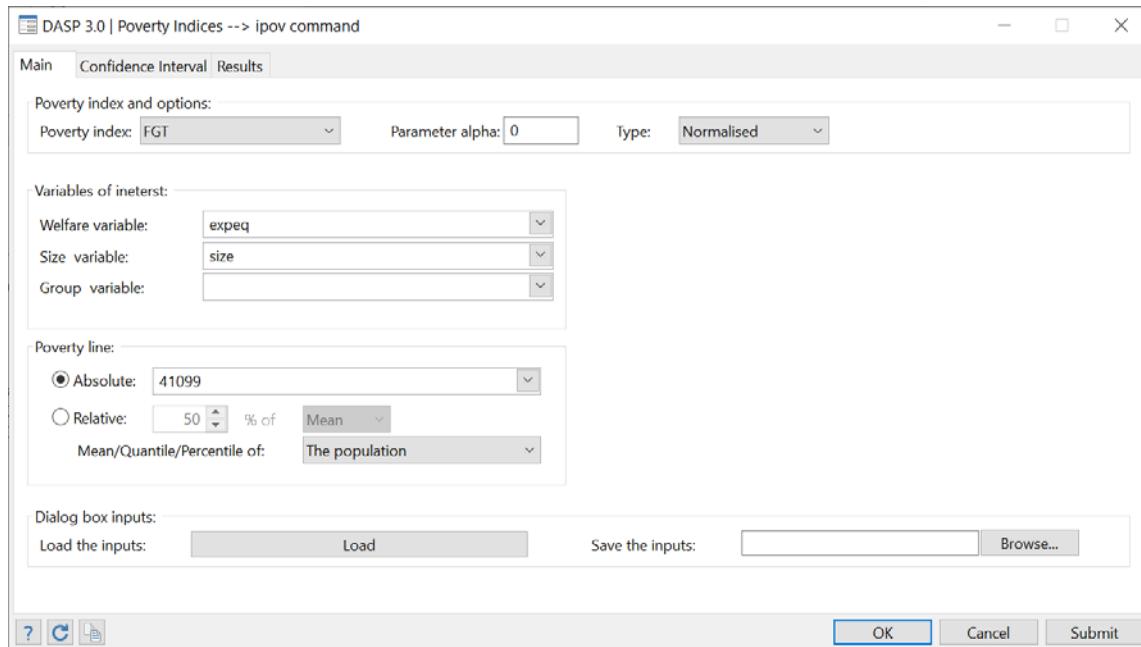
```
pweight: weight
          VCE: linearized
Single unit: missing
Strata 1: strata
SU 1: psu
FPC 1: <zero>
```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	42	838	19	20.0	20
2	37	733	17	19.8	20
3	98	1,959	19	20.0	20
4	55	1,093	19	19.9	20
5	66	1,286	13	19.5	20
6	41	779	1	19.0	20
7	97	1,937	19	20.0	20
	7	436	1	19.8	20
		8,625			

### Q.3

Type `bd ipov` to open the dialog box of poverty indices, and then choose variables and parameters as indicated in the following window. Click on **SUBMIT**.

Figure 26: Estimating FGT indices



The following results should then be displayed:

```
. ipov exppc expeq, alpha(0) hsize(size) index(fgt) pline(41099)
```

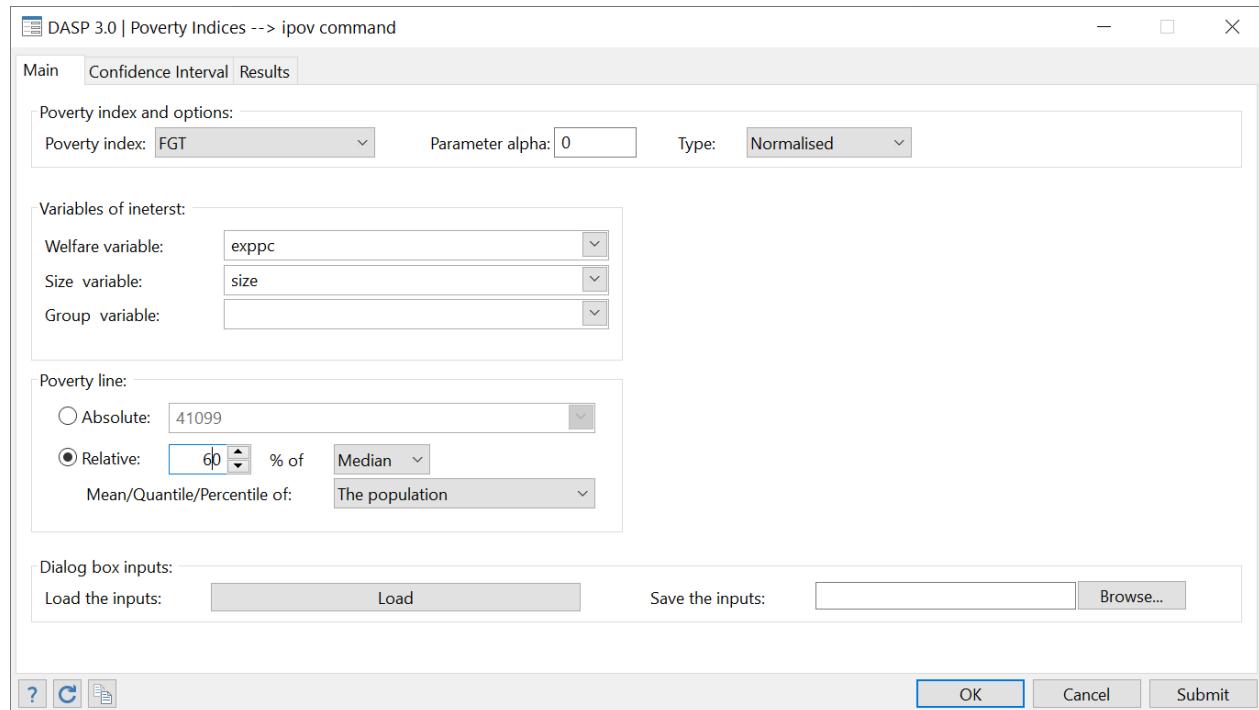
```
Poverty index : FGT index
Household size : size
Sampling weight : weight
Parameter alpha : 0.00
```

Variable	Estimate	STE	LB	UB	Pov. line
exppc	0.444565	0.016124	0.412873	0.476256	41099.00
expeq	0.255400	0.013326	0.229208	0.281592	41099.00

#### Q.4

Select RELATIVE for the poverty line and set the other parameters as above.

Figure 27: Estimating FGT indices with relative poverty lines



After clicking on SUBMIT, the following results should be displayed:

```
. ipov exppc, alpha(0) hsize(size) index(fgt) opl(median) prop(60)
```

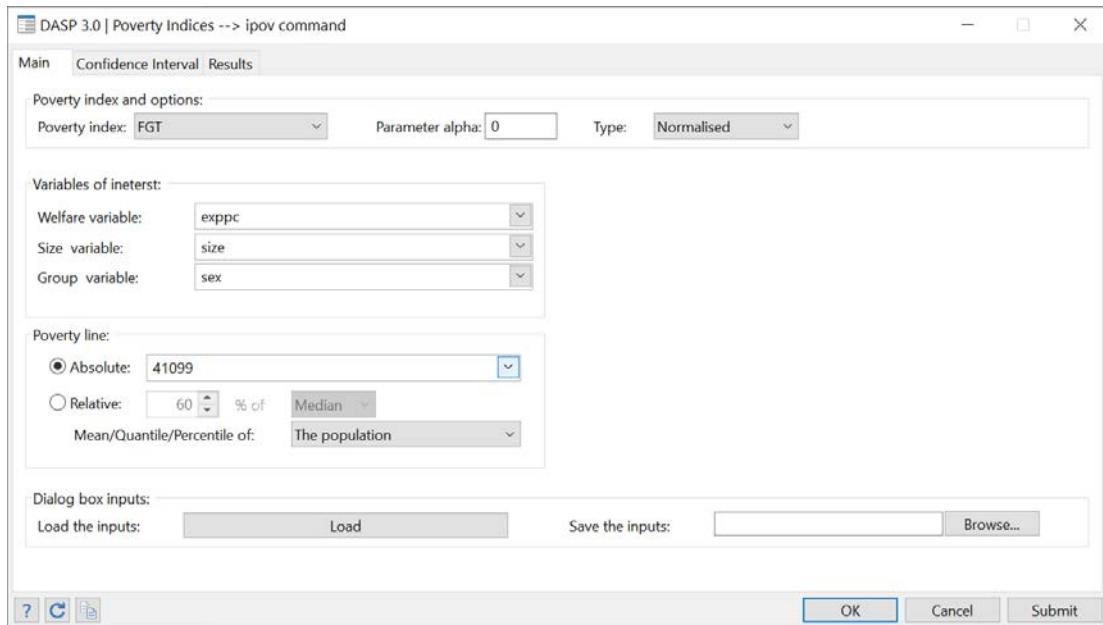
```
Poverty index : FGT index
Household size : size
Sampling weight : weight
Parameter alpha : 0.00
```

Variable	Estimate	STE	LB	UB	Pov. line
exppc	0.185243	0.008576	0.168386	0.202099	27046.71

## Q.5

Set the group variable to *sex*.

Figure 28: FGT indices differentiated by gender



Clicking on SUBMIT, the following should appear:

```
. ipov exppc, alpha(0) hsize(size) hgroup(sex) index(fgt) pline(41099)
```

```
Poverty index : FGT index
Household size : size
Sampling weight : weight
Group variable : sex
Parameter alpha : 0.00
```

Group	Estimate	STE	LB	UB	Pov. line
1: Male	0.452176	0.016633	0.419484	0.484867	41099.00
2: Female	0.281850	0.028206	0.226411	0.337290	41099.00
Population	0.444565	0.016124	0.412873	0.476256	41099.00

### Q.6

Using the panel CONFIDENCE INTERVAL, set the confidence level to 99 % and set the number of decimals to 4 in the RESULTS panel.

. ipov exppc, alpha(0) hsize(size) hgroup(sex) index(fgt) pline(41099) level(99)					
Poverty index : FGT index					
Household size : size					
Sampling weight : weight					
Group variable : sex					
Parameter alpha : 0.00					
Group	Estimate	STE	LB	UB	Pov. line
1: Male	0.452176	0.016633	0.409141	0.495210	41099.00
2: Female	0.281850	0.028206	0.208872	0.354829	41099.00
Population	0.444565	0.016124	0.402847	0.486282	41099.00

## 23.5 Estimating differences between FGT indices

"Has poverty Burkina Faso decreased between 1994 and 1998?"

1. Open the dialog box for the difference between poverty indices.
2. Estimate the difference between headcount indices when
  - a. Distribution 1 is year 1994 and distribution 2 is year 1998;
  - b. The variable of interest is **exppc** for 1994 and **exppcz** for 1998.
  - c. You should set size to household size in order to estimate poverty over the population of individuals.
  - d. Use 41099 Francs CFA per year as the poverty line for both distributions.
3. Estimate the difference between headcount indices when
  - a. Distribution 1 is rural residents in year 1998 and distribution 2 is rural residents in year 1994;
  - b. The variable of interest is **exppc** for 1994 and **exppcz** for 1998.
  - c. You should set size to household size in order to estimate poverty over the population of individuals.
  - d. Use 41099 Francs CFA per year as the poverty line for both distributions.
4. Redo the last exercise for urban residents.
5. Redo the last exercise only for members of male-headed households.
6. Test if the estimated difference in the last exercise is significantly different from zero. Thus, test:  
$$H_0: \Delta P(z = 41099, \alpha = 0) = 0 \text{ against } H_1: \Delta P(z = 41099, \alpha = 0) \neq 0$$
Set the significance level to 5% and assume that the test statistics follows a normal distribution.

## Answers

### Q.1

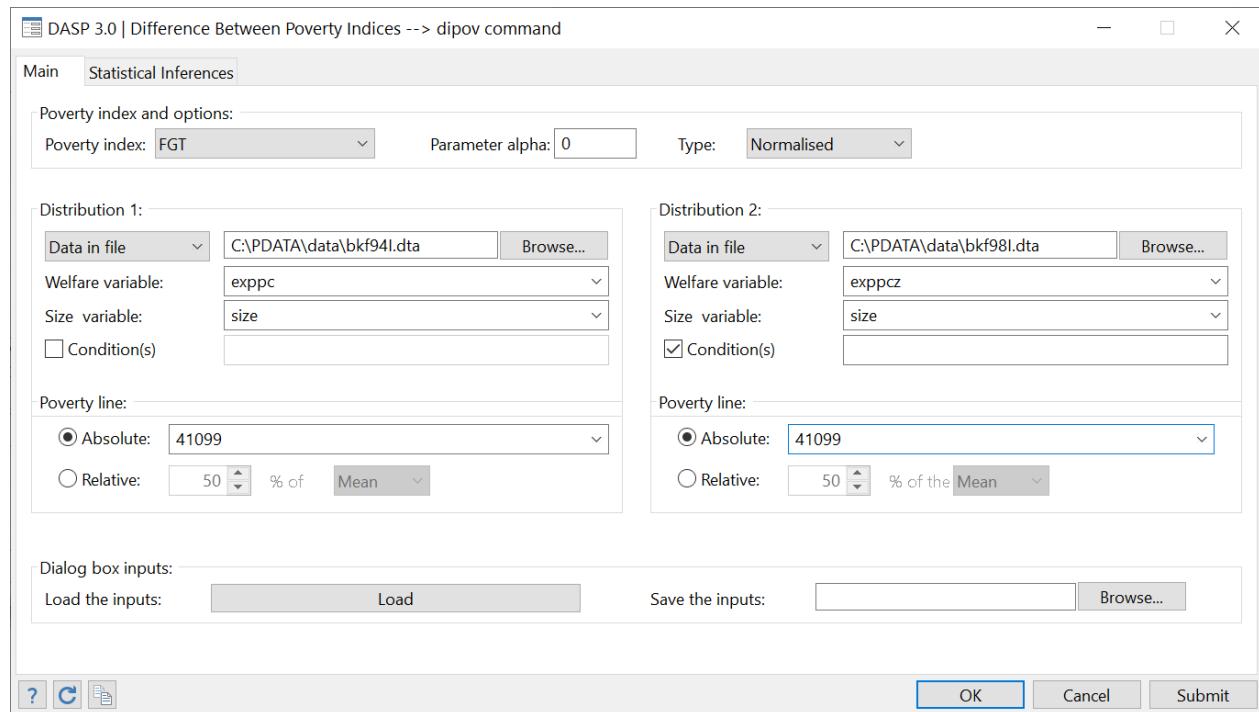
Open the dialog box by typing

*db dipov*

### Q.2

- For distribution 1, choose the option DATA IN FILE instead of DATA IN MEMORY and click on BROWSE to specify the location of the file **bkf98I.dta**.
- Follow the same procedure for distribution 2 to specify the location of **bkf94I.dta**.
- Choose variables and parameters as follows:

Figure 29: Estimating differences between FGT indices



After clicking on SUBMIT, the following should be displayed:

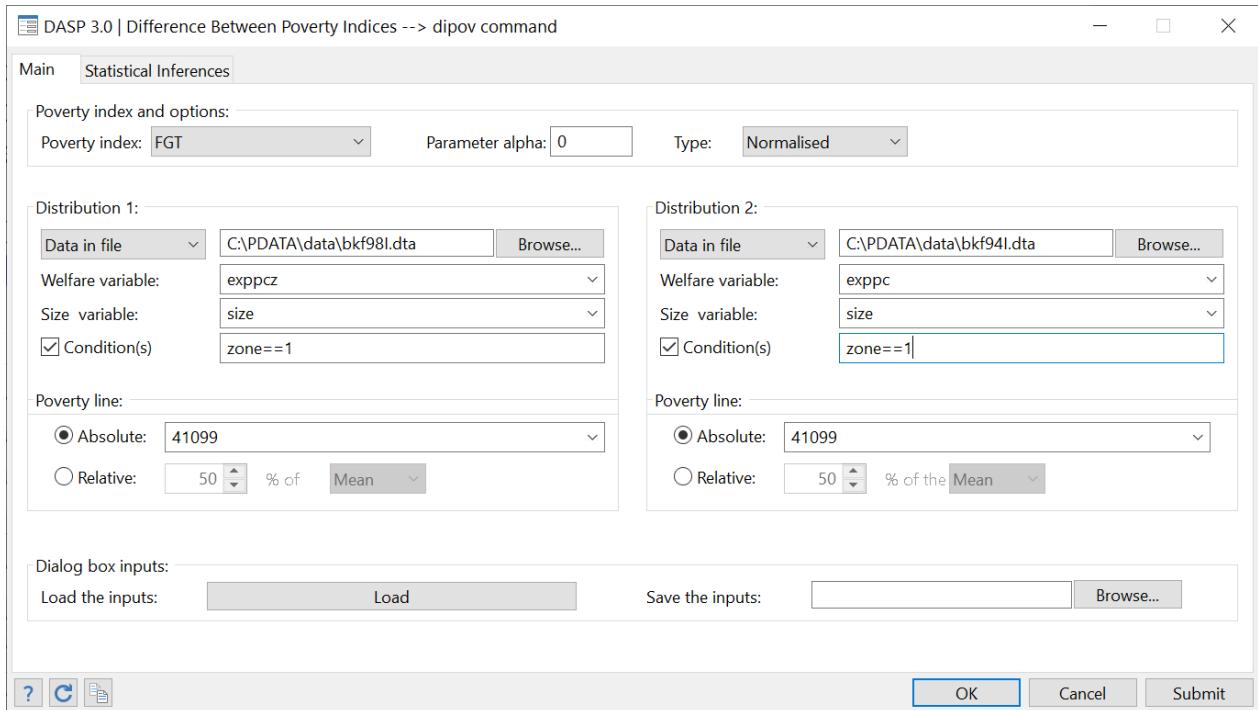
```
. dipov exppc exppcz, file1(C:\PDATA\data\bkf94I.dta) hsize1(size) file2(C:\PDATA\data\bkf98I.dta) hsize2(size) pline1(41099) pline2(41099) alpha(0) index(fgt)
Index : FGT index
```

Variable	Estimate	Std. Err.	t	P> t	[95% Conf. interval]	Pov. line
exppc	.4445646	.0161237	27.5721	0.0000	.4128733 .4762559	41099
exppcz	.4526775	.0109268	41.4282	0.0000	.4311987 .4741563	41099
diff.	.0081129	.0194774	.416529	0.6772	-.0301169 .0463427	---

### Q.3

- Restrict the estimation to rural residents as follows:
  - Select the option Condition(s)
  - Write ZONE in the field next to CONDITION (1) and type 1 in the next field.

Figure 30: Estimating differences in FGT indices



After clicking on SUBMIT, we should see:

Index : FGT index						
Variable	Estimate	Std. Err.	t	P> t	[95% Conf. interval]	Pov. line
exppc	.5104973	.0199749	25.5569	0.0000	.4712365	.5497581
exppcz	.5103441	.0116013	43.9903	0.0000	.4875395	.5331487
diff.	-.0001532	.0230995	-.006632	0.9947	-.0454924	.045186
						---

### Q.4

Index : FGT index						
Variable	Estimate	Std. Err.	t	P> t	[95% Conf. interval]	Pov. line
exppc	.1036838	.0134189	7.7267	0.0000	.0773088	.1300588
exppcz	.164573	.0162969	10.0984	0.0000	.1325382	.1966078
diff.	.0608892	.0211106	2.8843	0.0041	.0194538	.1023246
						---

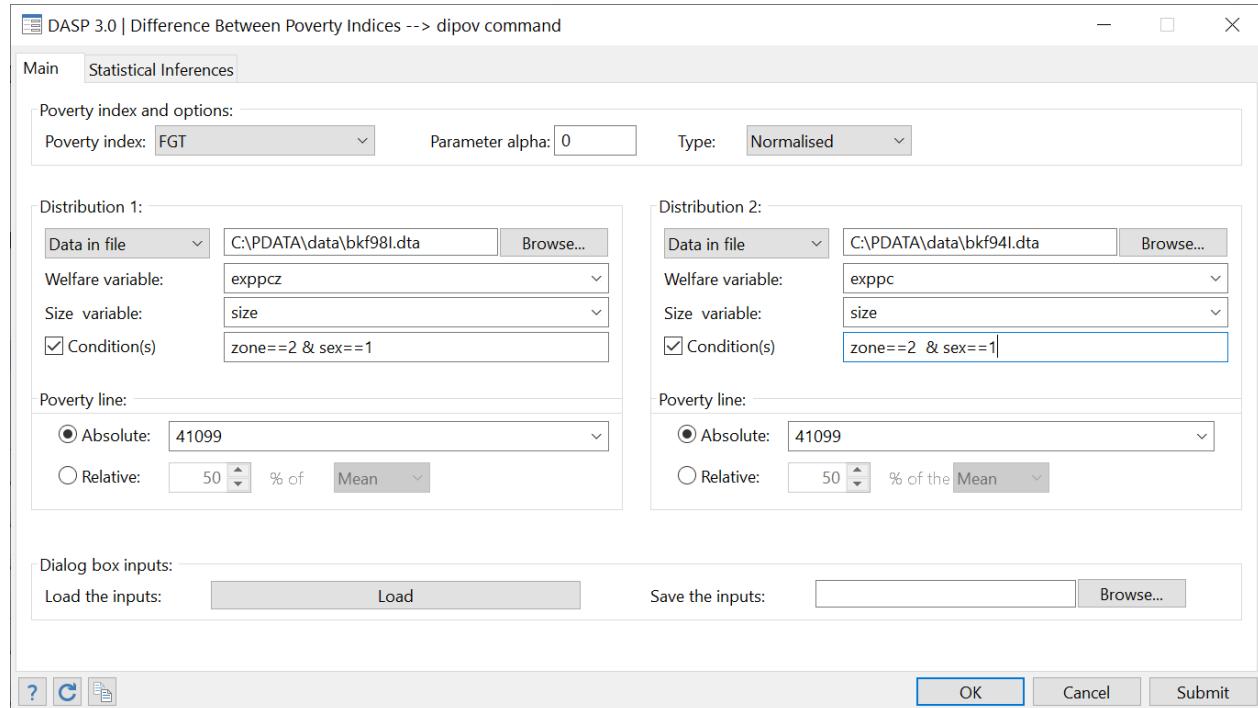
One can see that the change in poverty was significant only for urban residents.

### Q.5

Restrict the estimation to male-headed urban residents as follows:

- Set the number of Condition(s) to 2;
- Set **sex** in the field next to **Condition (2)** and type **1** in the next field.

Figure 31: FGT differences across years by gender and zone



After clicking on **SUBMIT**, the following should be displayed:

Index : FGT index

Variable	Estimate	Std. Err.	t	P> t	[95% Conf. interval]	Pov. line
exppc	.1059968	.0139448	7.60117	0.0000	.0785882	.1334054
exppcz	.1723844	.0177009	9.73874	0.0000	.1375898	.207179
diff.	.0663877	.022534	2.94611	0.0034	.0221584	.110617

### Q.6

We have that:

Lower Bound: = 0.0222

Upper Bound: = 0.1106

The null hypothesis is rejected since the lower bound of the 95% confidence interval is above zero.

## 23.6 Estimating multidimensional poverty indices

"How much is bi-dimensional poverty (total expenditures and literacy) in Peru in 1994?"

Using the **peru94I.dta** file,

1. Estimate the *Chakravarty et al. (1998)* index with parameter alpha = 1 and

	Var. of interest	Pov. line	a_j
Dimension 1	exppc	400	1
Dimension 2	pliterate	0.90	1

2. Estimate the *Bourguignon and Chakravarty (2003)* index with parameters alpha=beta=gamma= 1 and

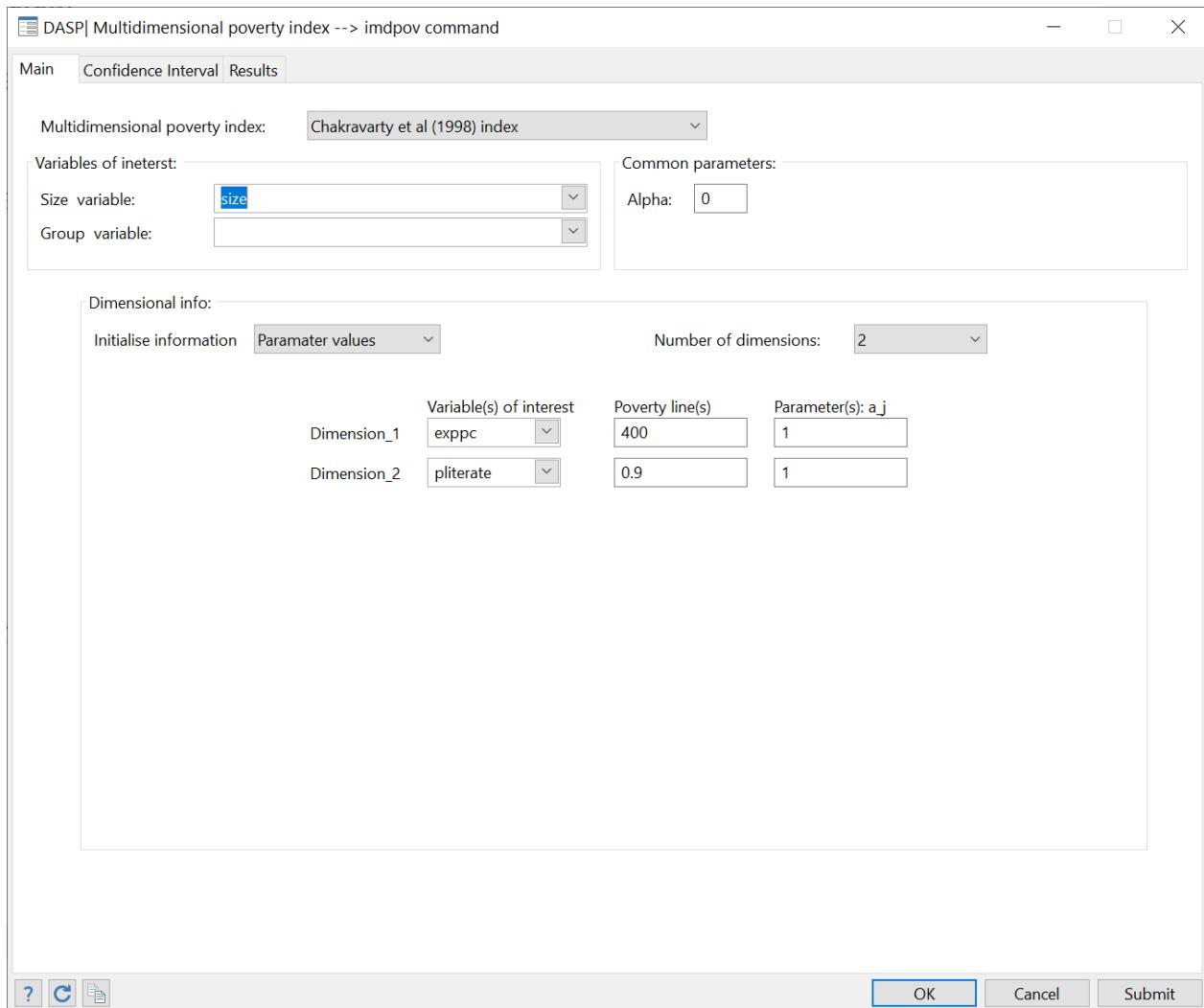
	Var. of interest	Pov. line
Dimension 1	exppc	400
Dimension 2	literate	0.90

### Q.1

Steps:

- Type  
*sysuse peru94\_A\_I.dta, clear*
- To open the relevant dialog box, type  
*db imdpov*
- Choose variables and parameters as in

Figure 32: Estimating multidimensional poverty indices (A)



After clicking **SUBMIT**, the following results appear.

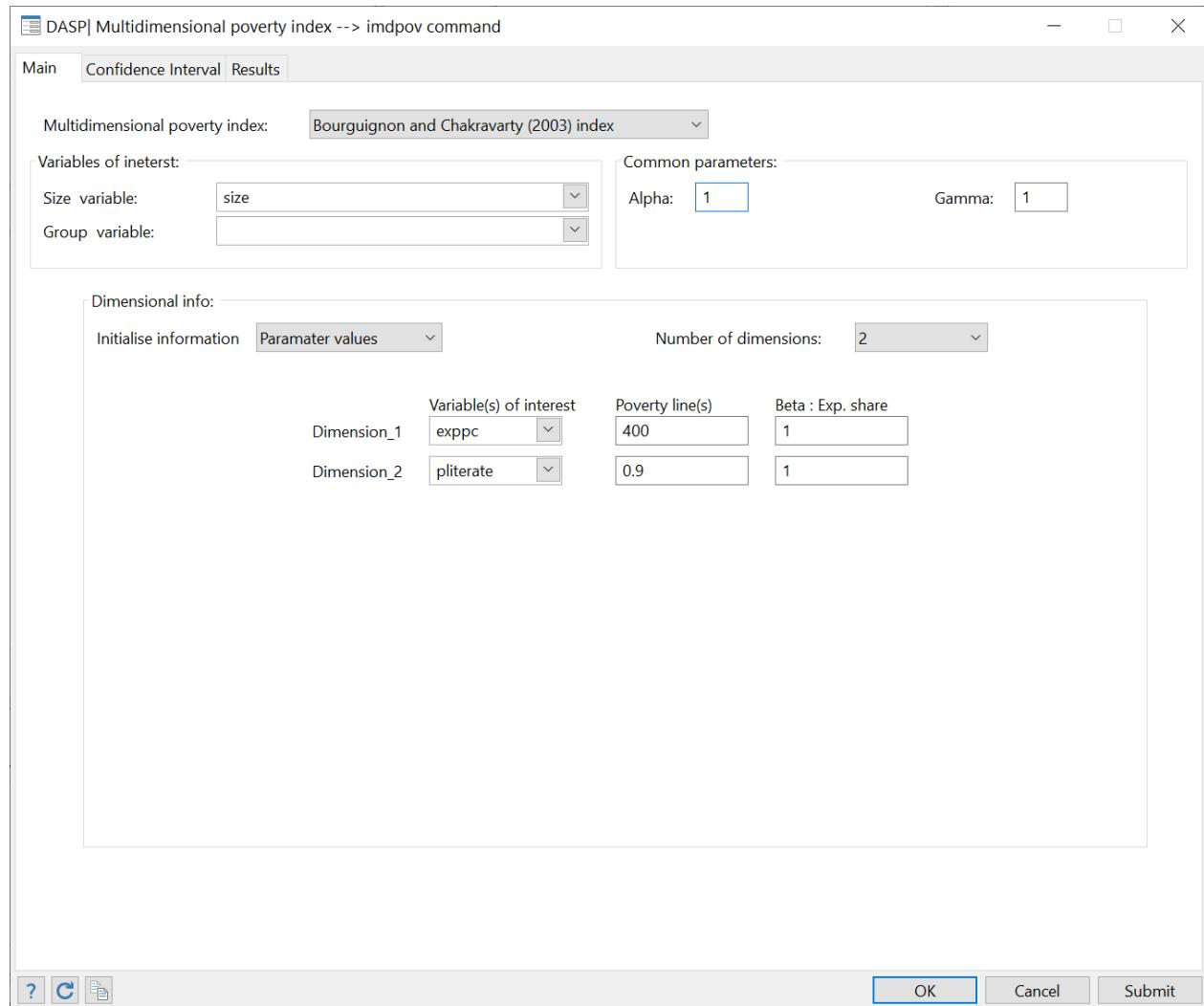
```
. imdpov exppc pliterate, index(1) pinfo(values) hsize(size) ndims(2) alpha(0) a1(1) p1(400) a2(1) p12(0.9)
```

```
M.D. Poverty index : Chakravarty et al (1998)
Household size      : size
```

	Estimate	STE	LB	UB
Population	0.418	0.009	0.403	0.433

## Q.2

Figure 33: Estimating multidimensional poverty indices (B)



After clicking SUBMIT, the following results appear.

```
. imdpov exppc pliterate, index(7) pinfo(values) hsize(size) ndims(2) alpha(1) gamma(1) b1(1) pl1(400) b2(1) pl2(0.9)
```

```
M.D. Poverty index : Bourguignon and Chakravarty (2003)
Household size      : size
```

	Estimate	STE	LB	UB
Population	0.098	0.003	0.093	0.103

## 23.7 Estimating FGT curves

"How sensitive to the choice of a poverty line is the rural-urban difference in poverty?"

1. Open ***bkf94l.dta***
2. Open the FGT curves dialog box.
3. Draw FGT curves for variables of interest *exppc* and *expeq* with
  - a. parameter  $\alpha = 0$ ;
  - b. poverty line between 0 and 100,000 Franc CFA;
  - c. size variable set to *size*;
  - d. subtitle of the figure set to "Burkina 1994".
4. Draw the 95% confidence interval FGT curves for urban and rural residents with
  - a. variable of interest set to *expcap*;
  - b. parameter  $\alpha = 0$ ;
  - c. poverty line between 0 and 100,000 Franc CFA;
  - d. size variable set to *size*.
5. Draw the difference between these two curves and
  - a. save the graph in \*.gph format to be plotted in Stata and in \*.wmf format to be inserted in a Word document.
  - b. List the coordinates of the graph.
6. Redo the last graph with  $\alpha = 1$ .

### Answers

#### Q.1

Open the file with

```
sysuse b kf94l.dta, clear
```

#### Q.2

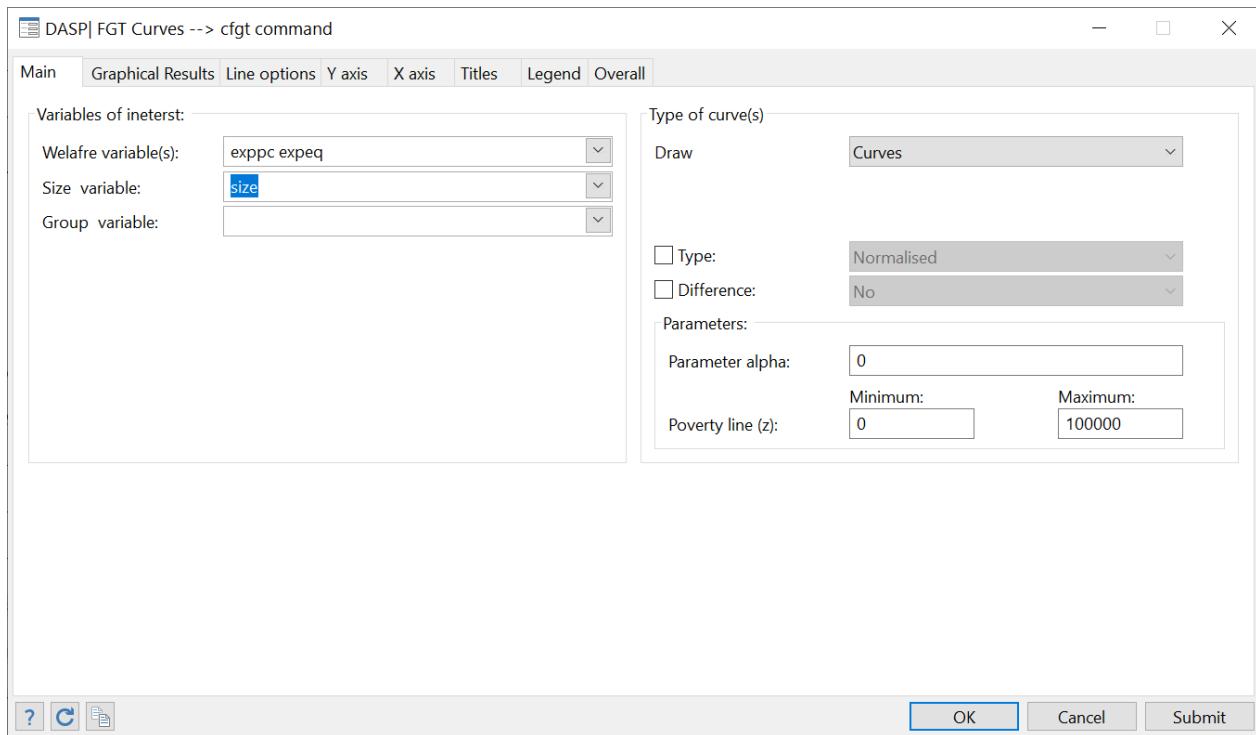
Open the dialog box by typing

```
db cfgt
```

#### Q.3

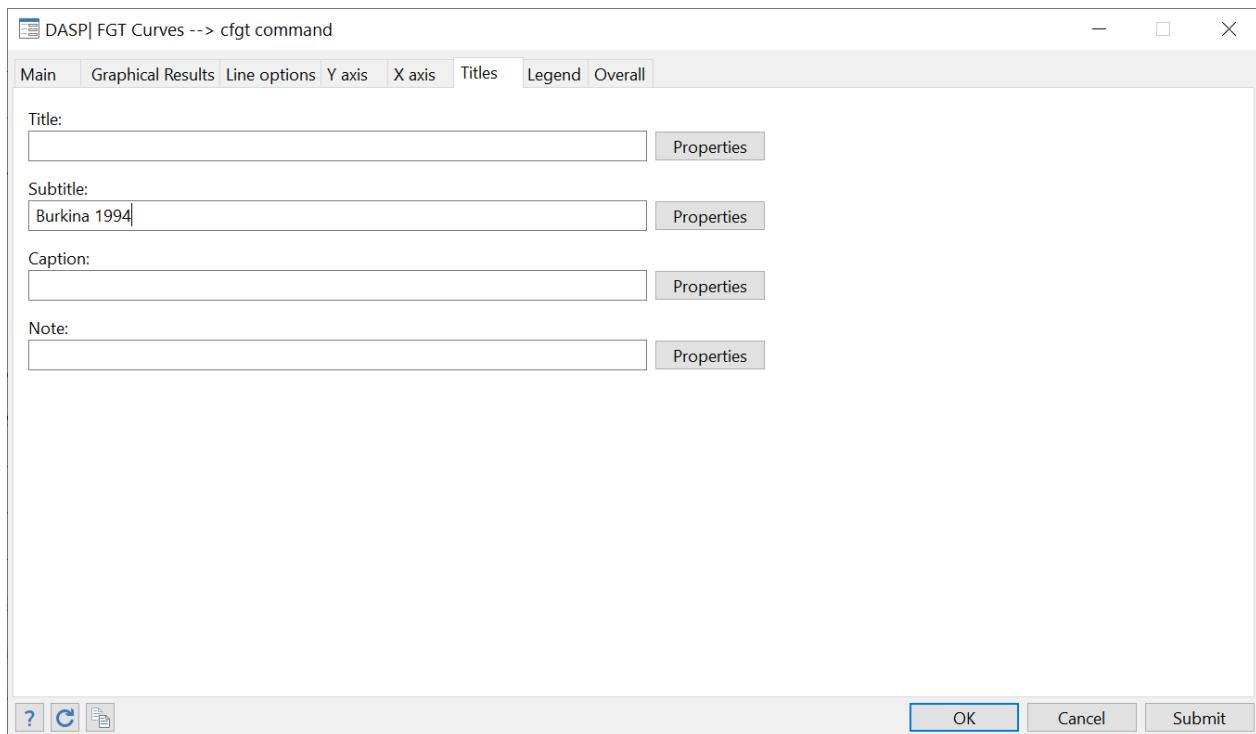
Choose variables and parameters as follows:

**Figure 34: Drawing FGT curves**



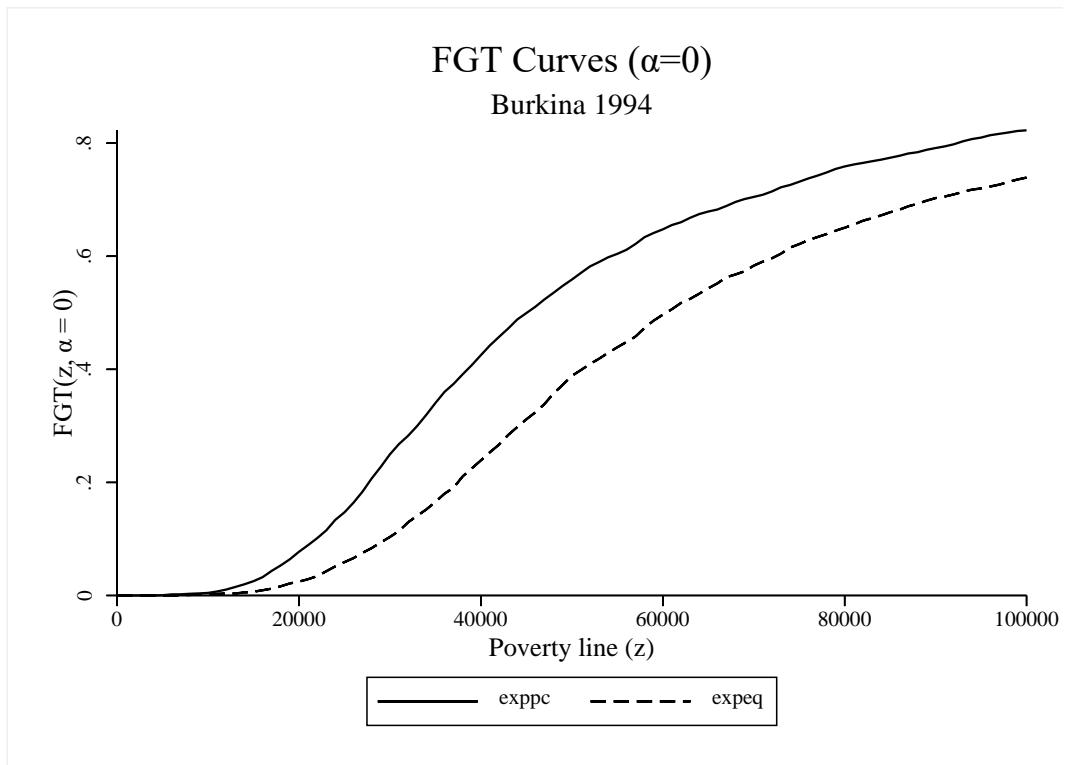
To change the subtitle, select the Title panel and write the subtitle.

**Figure 35: Editing FGT curves**



After clicking SUBMIT, the following graph appears:

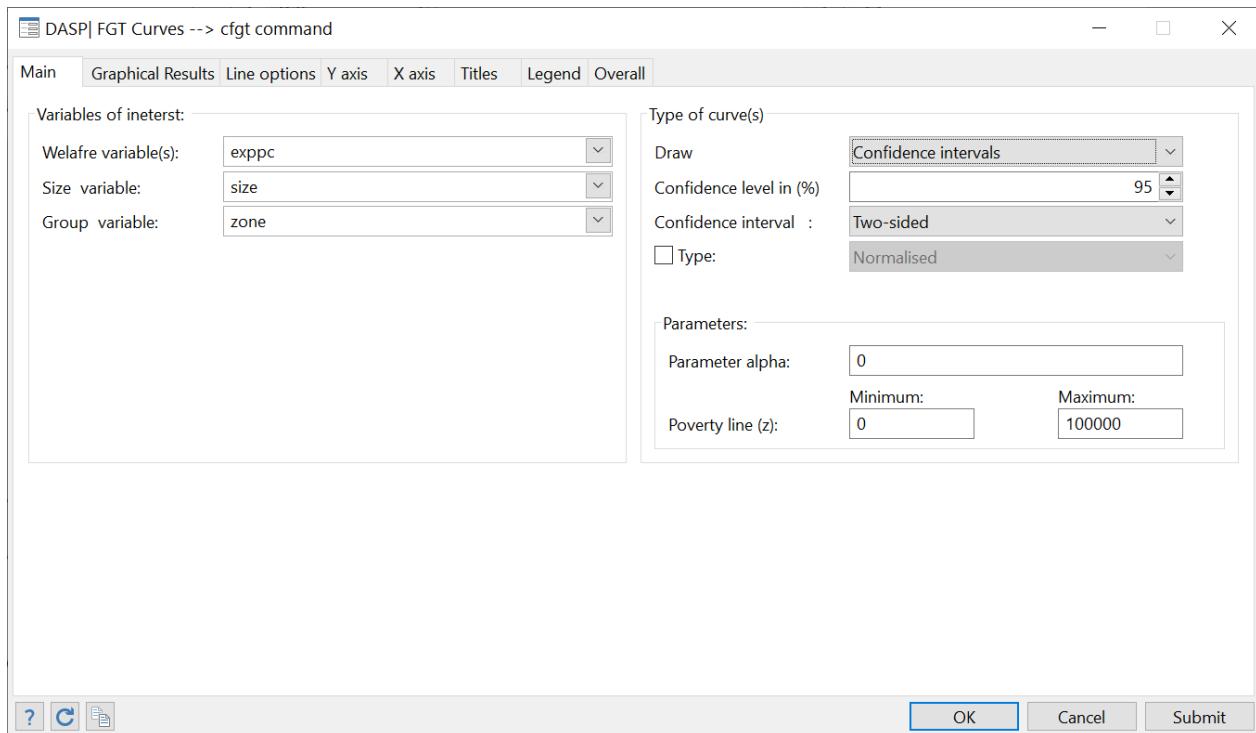
Figure 36: Graph of FGT curves



#### Q.4

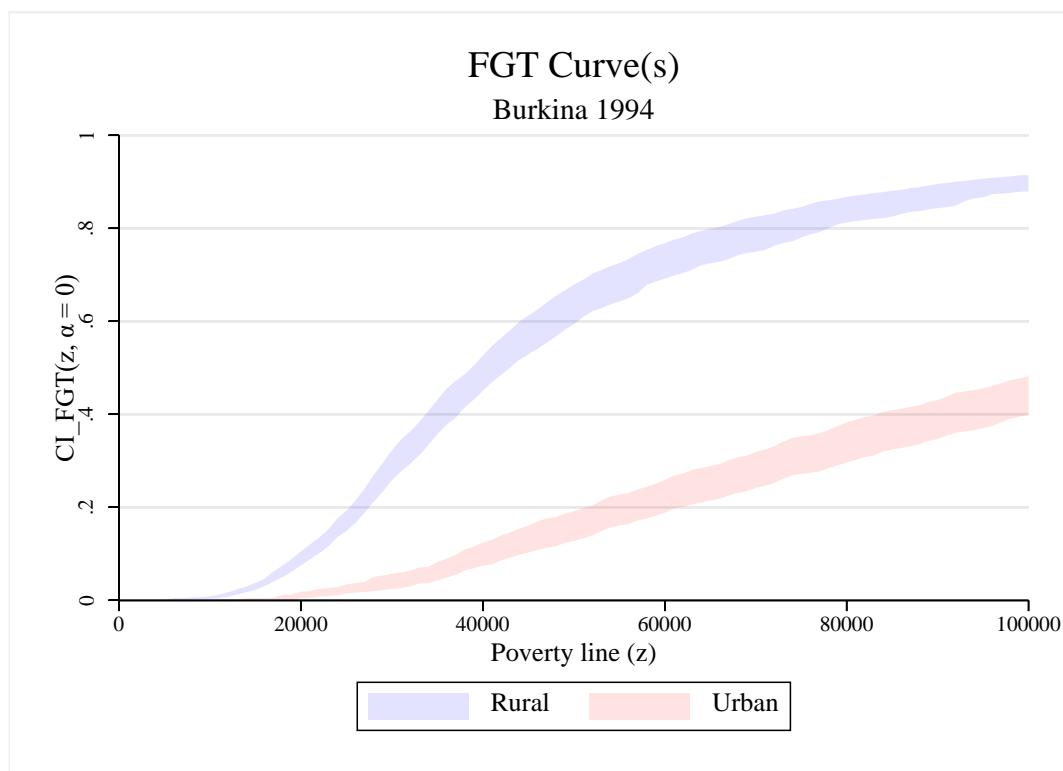
Choose variables and parameters as in the following window:

Figure 37: FGT curves by zone



After clicking **SUBMIT**, the following graph appears:

Figure 38: Graph of FGT curves by zone



### Q.5

- Choose the option DIFFERENCE and select: WITH THE FIRST CURVE;
- Indicate that the group variable is *zone*;
- Select the Results panel and choose the option LIST in the COORDINATES quadrant.
- In the GRAPH quadrant, select the directory in which to save the graph in gph format and to export the graph in wmf format.

Figure 39: Differences of FGT curves

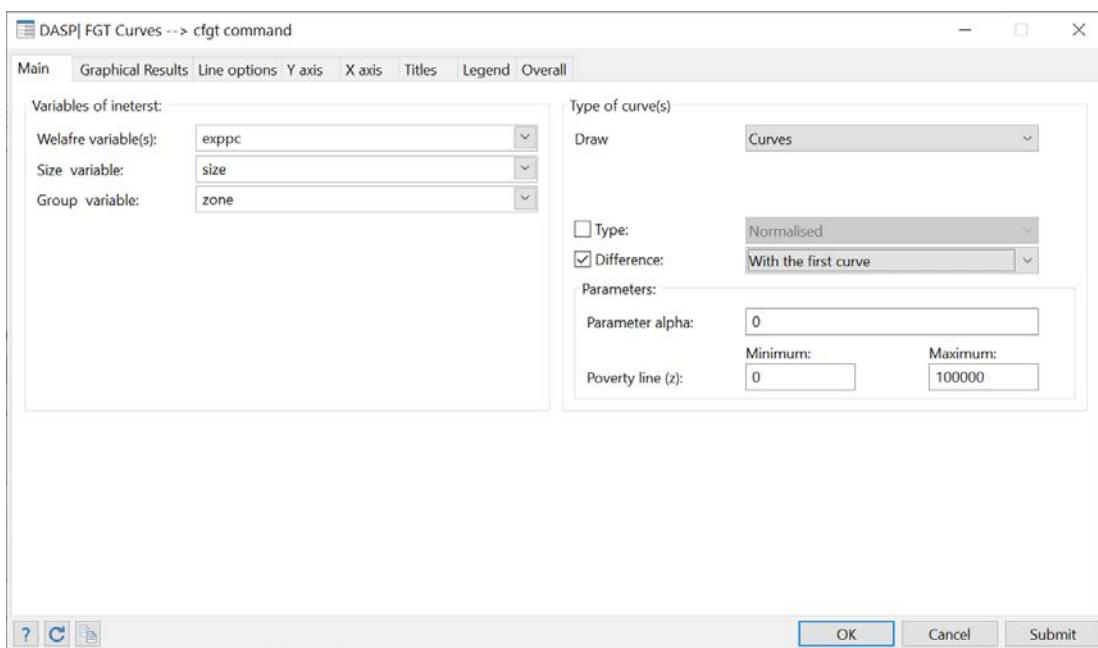
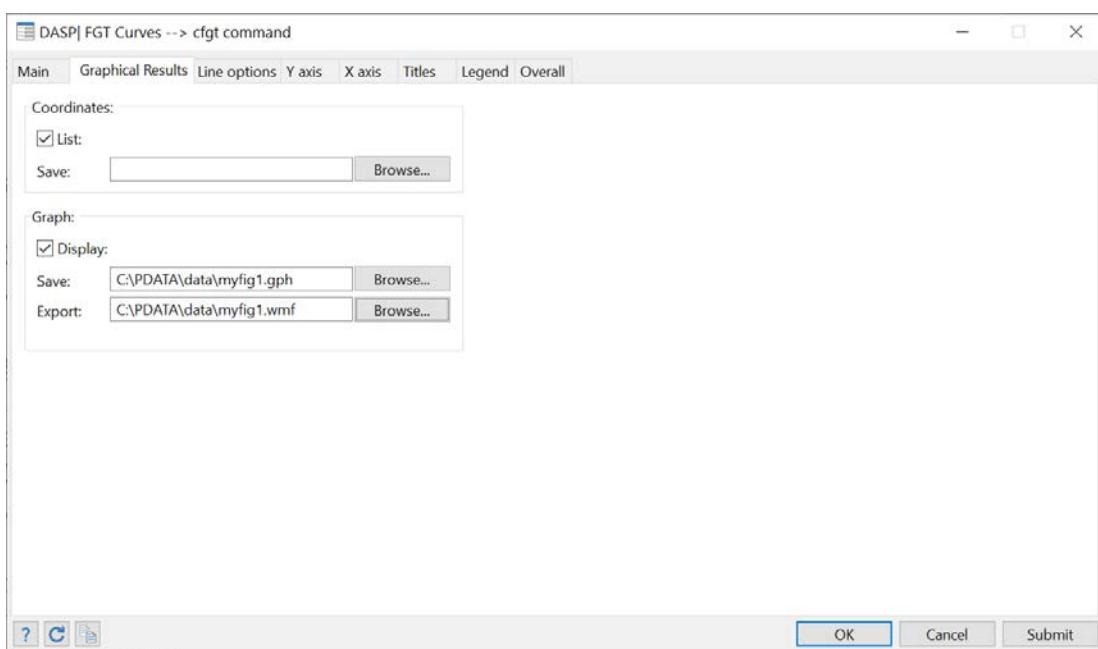
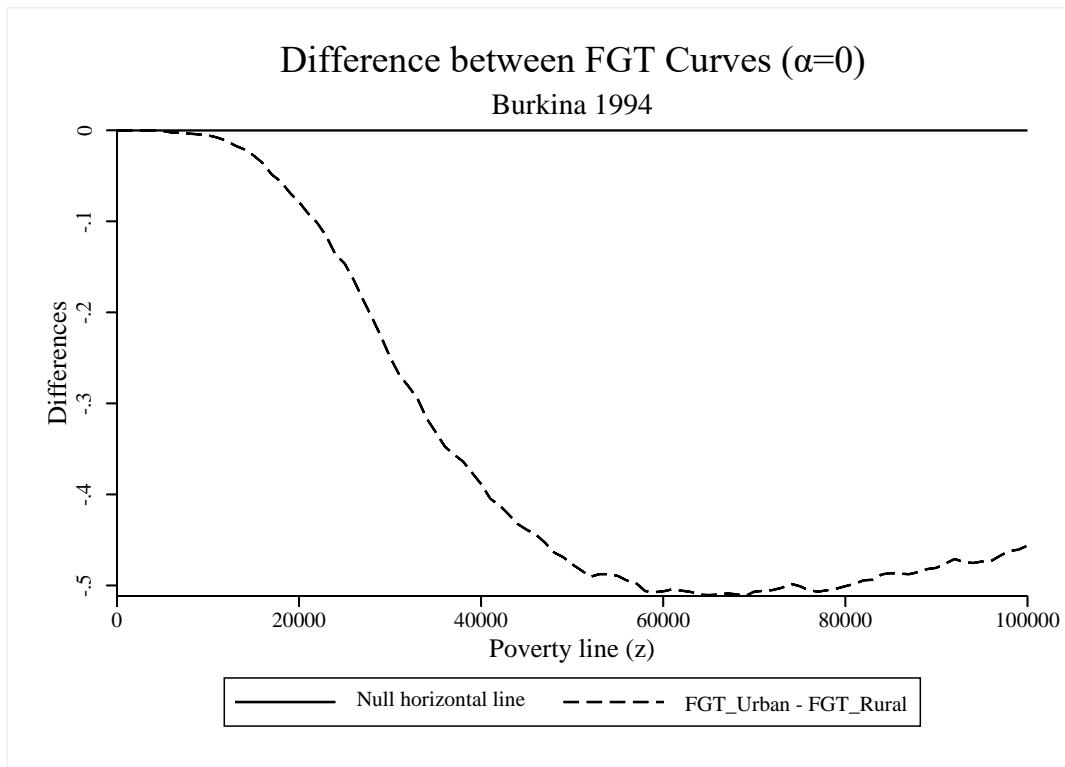


Figure 40: Listing coordinates



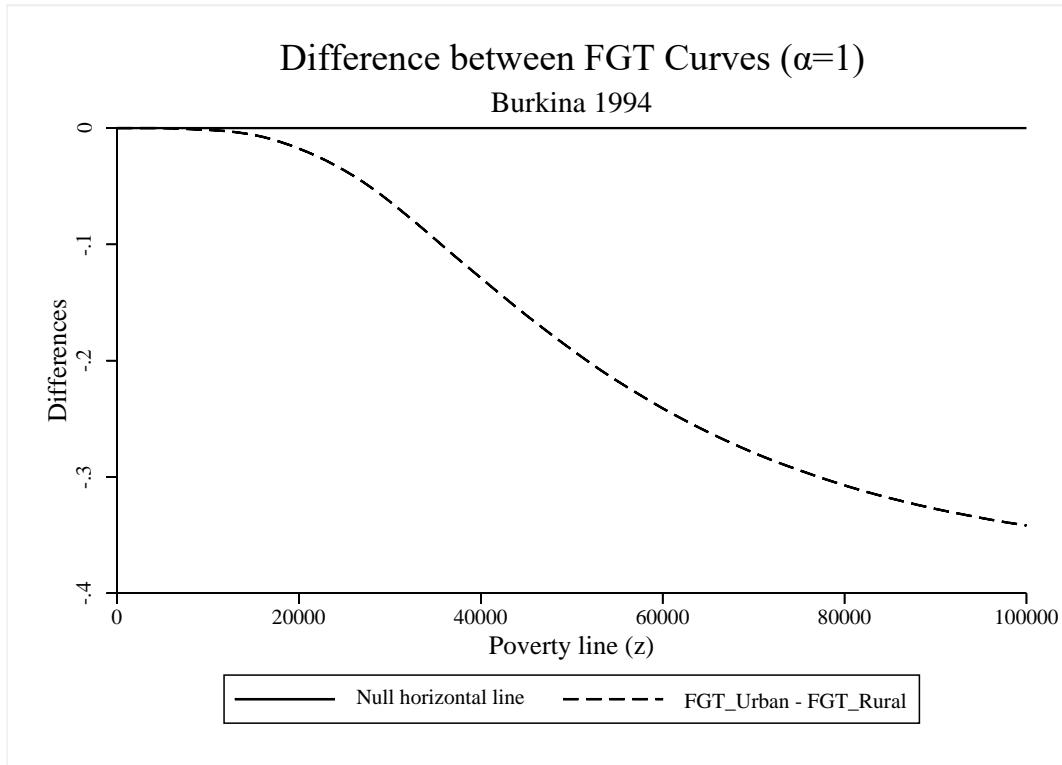
After clicking SUBMIT, the following appears:

Figure 41: Differences between FGT curves



## Q.6

Figure 42: Differences between FGT curves



### 23.8 Estimating differences between FGT curves with confidence intervals

"Is the poverty increase between 1994 and 1998 in Burkina Faso statistically significant?"

- 1) Using simultaneously the files **bkf94I.dta** and **bkf98I.dta**, draw the difference between FGT curves and associated confidence intervals with:
  - a) The variable of interest *exppc* for 1994 and *exppcz* for 1998.
  - b) parameter  $\alpha = 0$ ;
  - c) poverty line between 0 and 100,000 Franc CFA;
  - d) size variable set to *size*.
- 2) Redo 2) with parameter  $\alpha = 1$ .

## Answers

### Q.1

Steps:

- To open the relevant dialog box, type  
*db dicfgt*
- Choose variables and parameters as in

Figure 43: Drawing the difference between FGT curves with confidence interval

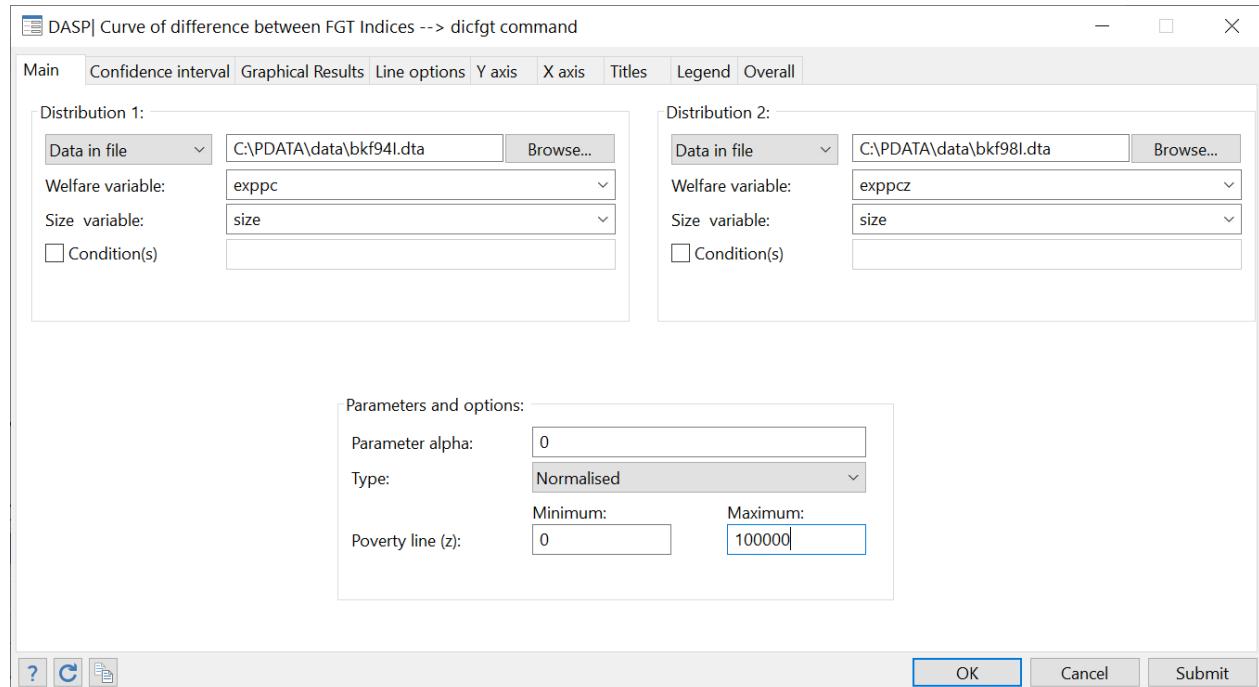


Figure 44: Difference between FGT curves with confidence interval ( $\alpha = 0$ )

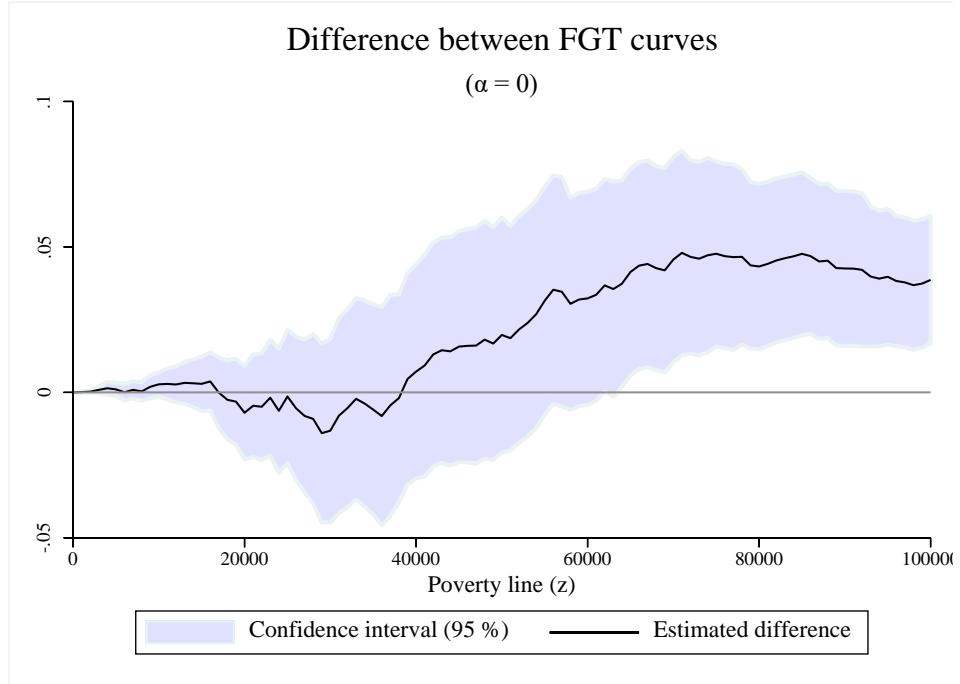
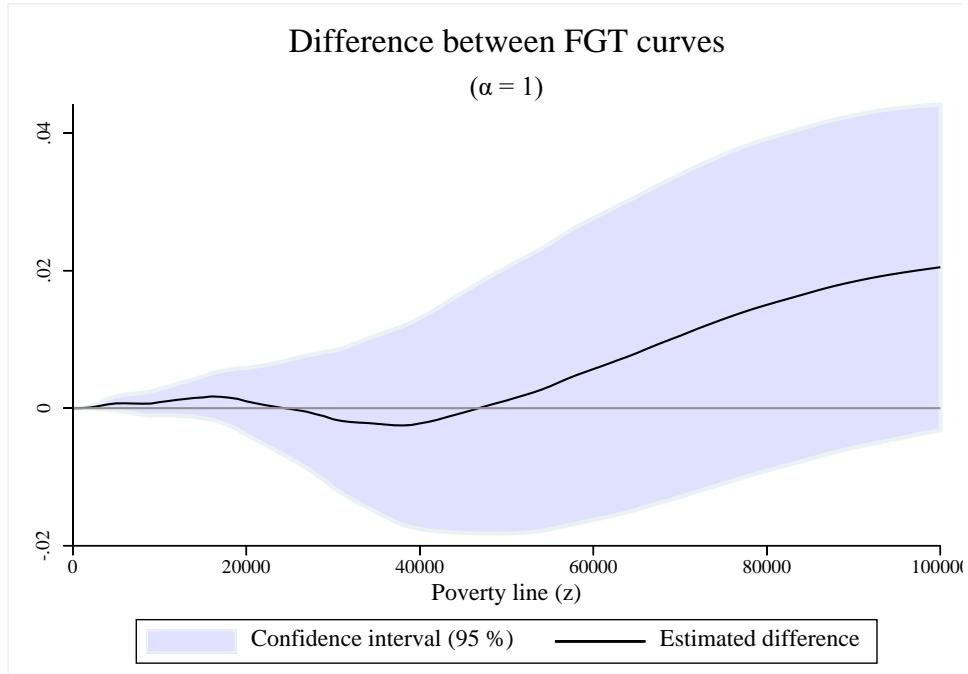


Figure 45: Difference between FGT curves with confidence interval ( $\alpha = 1$ )



## 23.9 Testing poverty dominance and estimating critical values.

"Has the poverty increase in Burkina Faso between 1994 and 1998 been statistically significant?"

- 1) Using simultaneously files *bkf94I.dta* and *bkf98I.dta*, check for second-order poverty dominance and estimate the values of the poverty line at which the two FGT curves cross.
  - a) The variable of interest is *exppc* for 1994 and *exppcz* for 1998;
  - b) The poverty line should vary between 0 and 100,000 Franc CFA;
  - c) The size variable should be set to *size*.

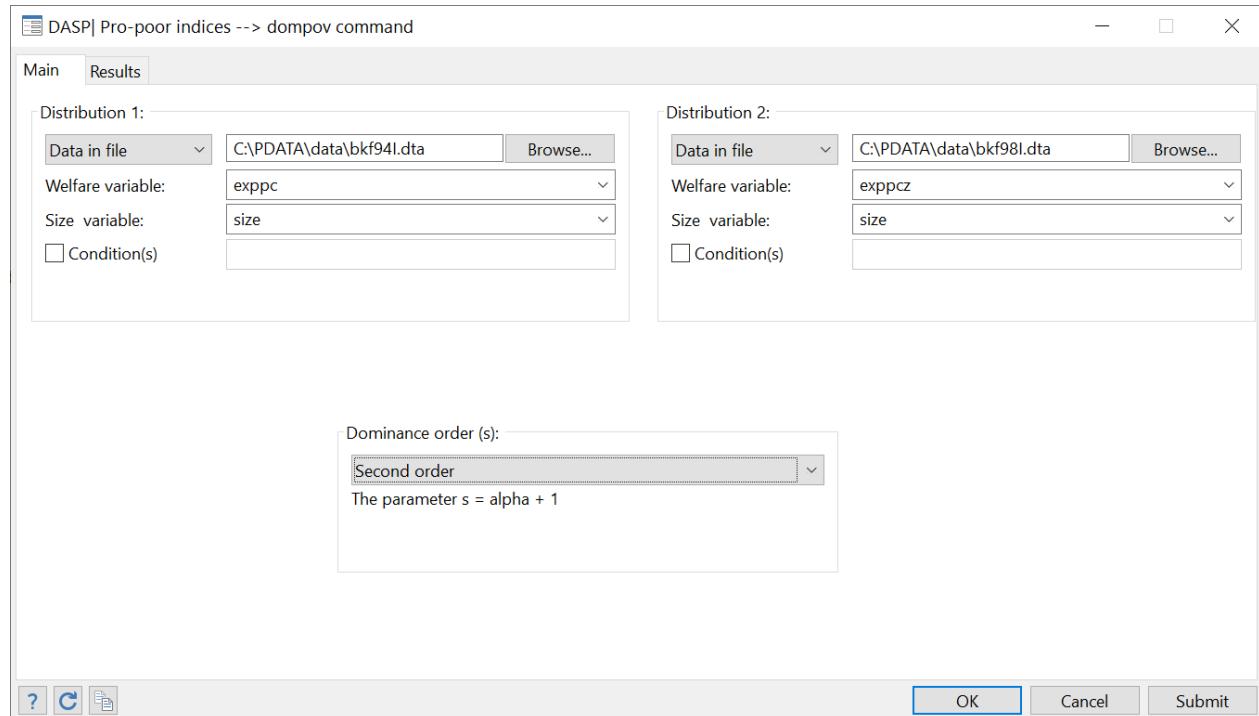
## Answers

### Q.1

Steps:

- To open the relevant dialog box, type  
*db dompov*
- Choose variables and parameters as in

Figure 46: Testing for poverty dominance



After clicking **SUBMIT**, the following results appear:

```
. dompov exppc exppcz, file1(C:\PDATA\data\bkf94I.dta) hsize1(size) file2(C:\PDATA\data\bkf98I.dta) hsize2(size) order(2)
```

Number of intersection	Critical pov. line	Min. range of pov. lines	Max. range of pov. line	Case
1	24262.871	.	.	A
2	46775.652	.	.	B

Notes :  
\_case A: Before this intersection, distribution 1 dominates (in welfare) distribution 2.  
\_case B: Before this intersection, distribution 2 dominates (in welfare) distribution 1.  
\_case C: No dominance before this intersection.

## 23.10 Decomposing FGT indices

"What is the contribution of different types of earners to total poverty in Burkina Faso?"

1. Open **bkf94I.dta** and decompose the average poverty gap
  - a. with variable of interest **exppc**;
  - b. with size variable set to **size**;
  - c. at the official poverty line of 41099 Francs CFA;
  - d. and using the group variable **gse** (socio-economic groups).
2. Do the above exercise without standard errors and with the number of decimals set to 4.

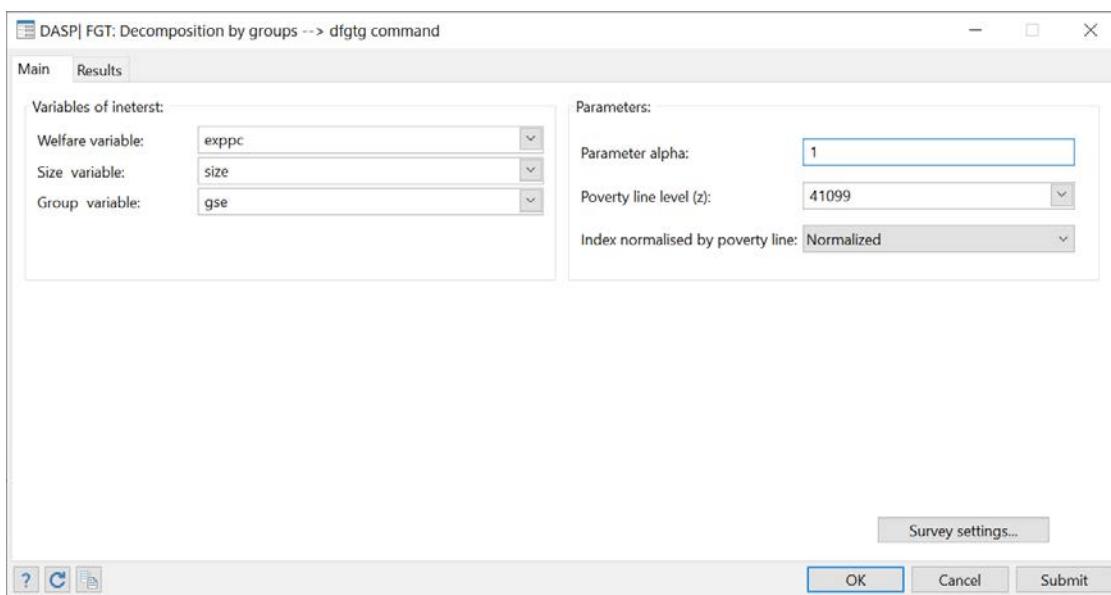
## Answers

### Q.1

Steps:

- Type
- **sysuse b kf94I.dta, clear**
  
- To open the relevant dialog box, type  
**db dfgtg**
- Choose variables and parameters as in

**Figure 47: Decomposing FGT indices by groups**



After clicking **SUBMIT**, the following information is provided:

```
. dfgtg exppc, hgroup(gse) hsize(size) alpha(1) pline(41099)
```

```
Decomposition of the FGT index by groups
Poverty index : FGT index
Household size : size
Sampling weight : weight
Group variable : gse
Parameter alpha : 1.00
```

Group	FGT index	Population share	Absolute contribution	Relative contribution
wage-earner (public sector)	0.004237 0.002571	0.042971 0.003790	0.000182 0.000117	0.001308 0.000840
wage-earner (private sector)	0.022176 0.010678	0.026598 0.002164	0.000590 0.000291	0.004237 0.002083
Artisan or trader	0.027741 0.004653	0.062640 0.004288	0.001738 0.000325	0.012484 0.002371
Other type of earner	0.063853 0.025805	0.066550 0.001308	0.000425 0.000170	0.003050 0.001203
Crop farmer	0.137525 0.011808	0.104402 0.014896	0.014358 0.002459	0.103148 0.016980
Subsistence farmer	0.162894 0.008643	0.680885 0.016403	0.110912 0.005823	0.796800 0.019015
Inactive	0.144916 0.014994	0.075856 0.004839	0.010993 0.001332	0.078973 0.008520
Population	0.139197 0.006553	1.000000 0.000000	0.139197 0.006553	1.000000 0.000000

## Q.2

Using the RESULTS panel, change the number of decimals and deselect the option DISPLAY STANDARD ERRORS.

After clicking SUBMIT, the following information is obtained:

```
. dfgtg exppc, hgroup(gse) hsize(size) alpha(1) pline(41099) dste(0) dec(4)
```

```
Decomposition of the FGT index by groups
Poverty index      : FGT index
Household size     : size
Sampling weight    : weight
Group variable     : gse
Parameter alpha    : 1.00
```

Group	FGT index	Population share	Absolute contribution	Relative contribution
wage-earner (public sector)	0.0042	0.0430	0.0002	0.0013
wage-earner (private sector)	0.0222	0.0266	0.0006	0.0042
Artisan or trader	0.0277	0.0626	0.0017	0.0125
Other type of earner	0.0639	0.0066	0.0004	0.0031
Crop farmer	0.1375	0.1044	0.0144	0.1031
Subsistence farmer	0.1629	0.6809	0.1109	0.7968
Inactive	0.1449	0.0759	0.0110	0.0790
Population	0.1392	1.0000	0.1392	1.0000

## 23.11 Estimating Lorenz and concentration curves

"How much do taxes and transfers affect inequality in Canada?"

By using the **can6.dta** file,

1. Draw the Lorenz curves for gross income  $X$  and net income  $N$ . How can you see the redistribution of income?
2. Draw Lorenz curves for gross income  $X$  and concentration curves for each of the three transfers  $B_1$ ,  $B_2$  and  $B_3$  and the tax  $T$ . What can you say about the progressivity of these elements of the tax and transfer system?

"What is the extent of inequality among Burkina Faso rural and urban households in 1994?"

By using the **bkf94I.dta** file,

3. Draw Lorenz curves for rural and urban households

- a. with variable of interest **exppc**;

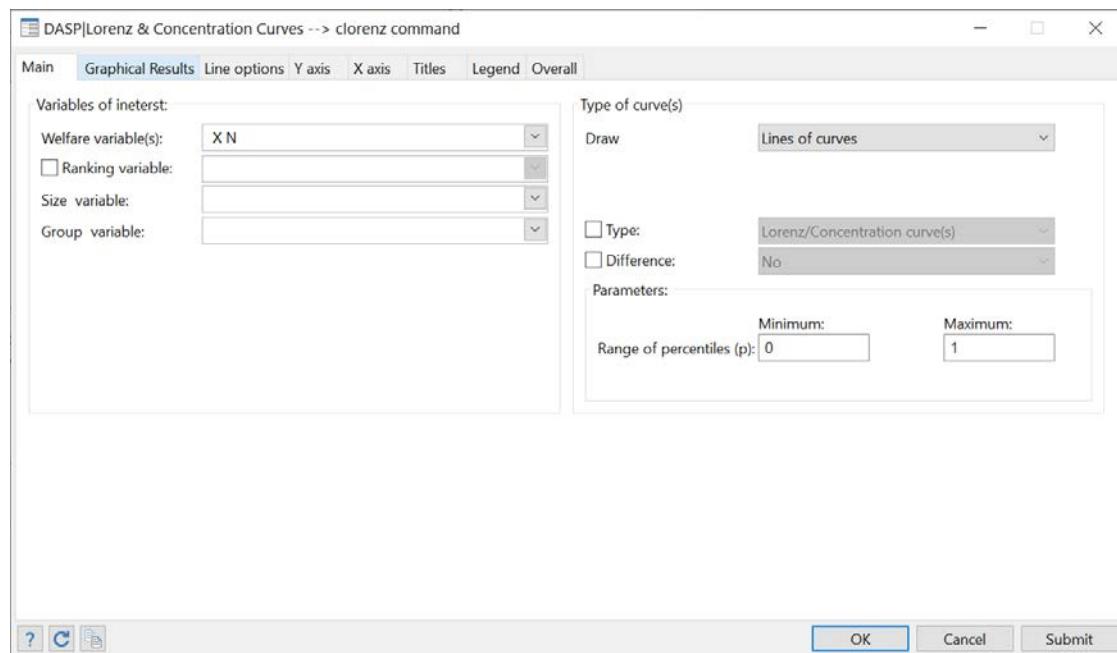
- b. with size variable set to `size`;
- c. and using the group variable `zone` (as residential area).

## Q.1

Steps:

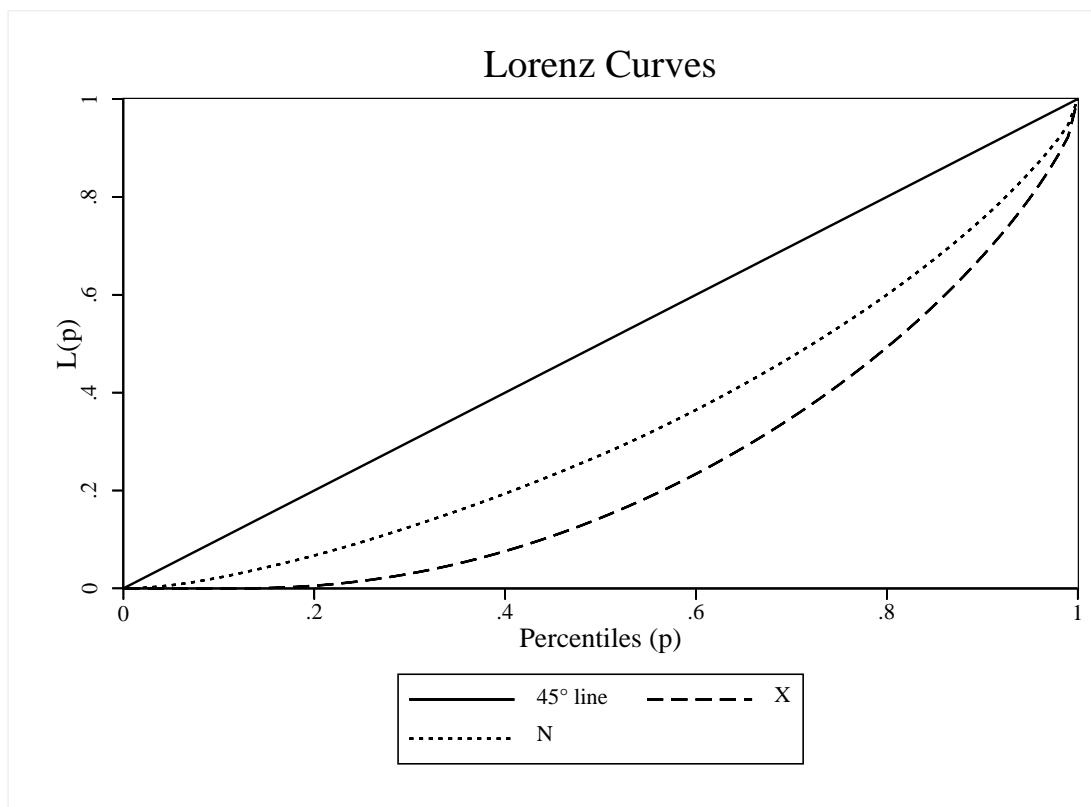
- Type  
`sysuse can6.dta, clear`
- To open the relevant dialog box, type  
`db clorenz`
- Choose variables and parameters as in

**Figure 48: Lorenz and concentration curves**



After clicking SUBMIT, the following appears:

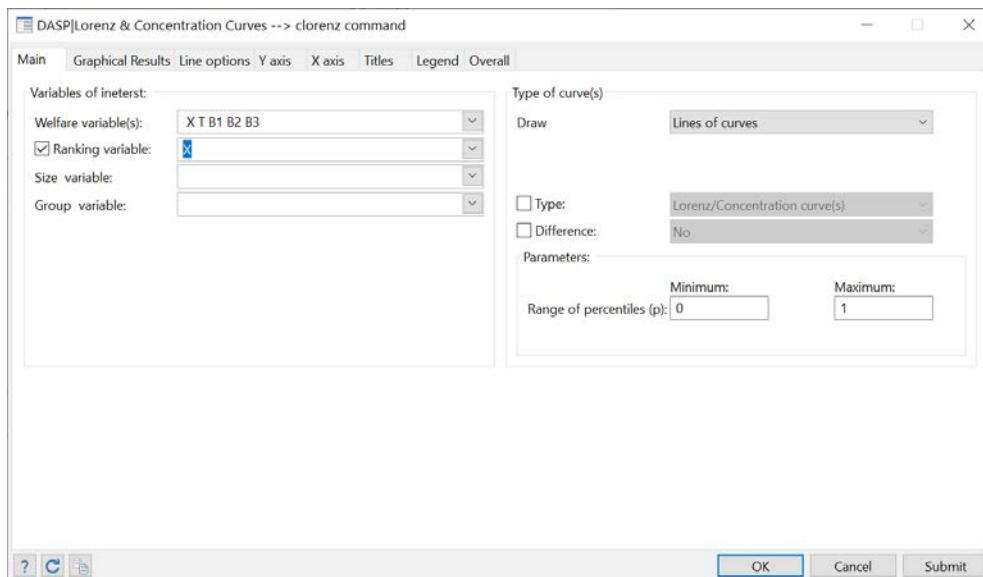
Figure 49: Lorenz curves



### Q.2

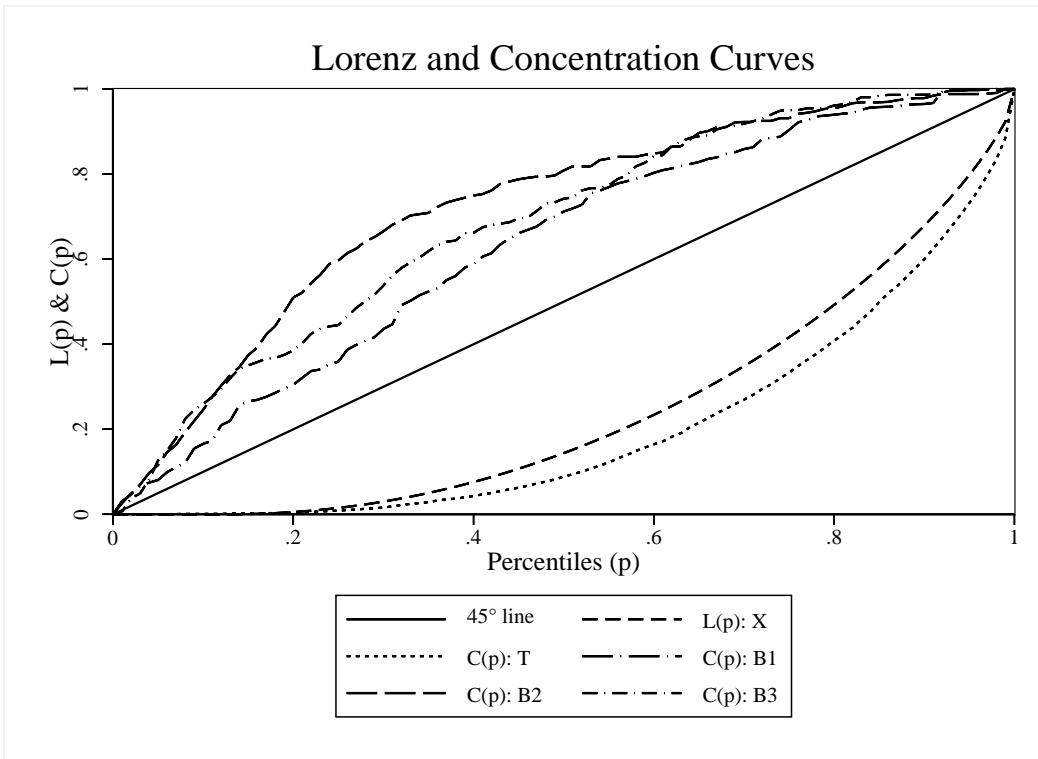
Steps: Choose variables and parameters as in the following dialog box:

Figure 50: Drawing concentration curves



After clicking on **SUBMIT**, the following appears:

Figure 51: Lorenz and concentration curves

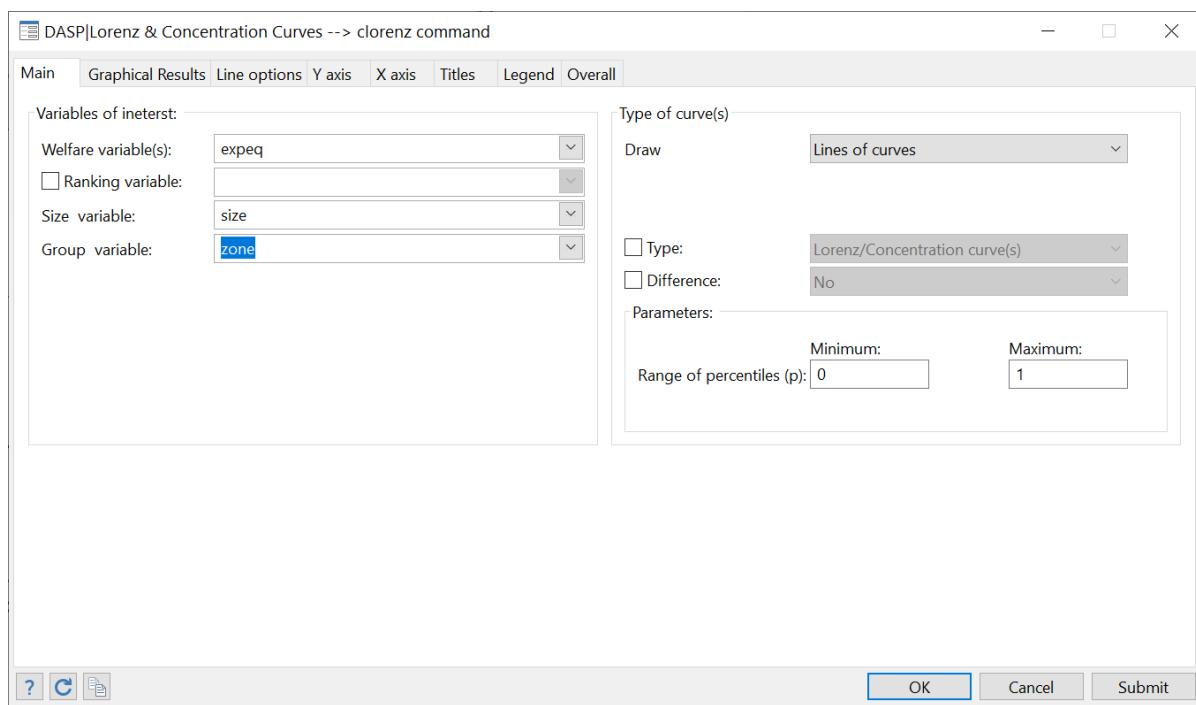


### Q.3

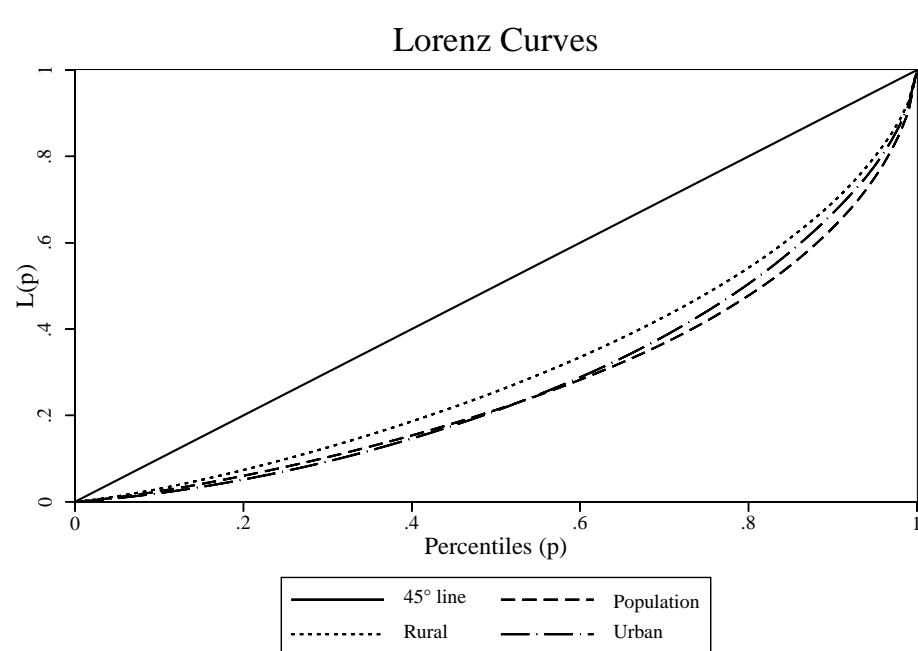
Steps:

- Type: `sysuse bkf94l.dta, clear`
- Choose variables and parameters as in the following dialog box:

**Figure 52 :Drawing Lorenz curves**



**Figure 53: Lorenz curves**



## 23.12 Estimating Gini and concentration curves

"By how much taxes and transfers affect inequality in Canada?"

Using the ***can6.dta*** file,

1. Estimate the Gini indices for gross income  $X$  and net income  $N$ .
2. Estimate the concentration indices for variables  $T$  and  $N$  when the ranking variable is gross income  $X$ .

"By how much inequality has changed in Burkina Faso between 1994 and 1998?"

Using the ***bkf94I.dta*** file,

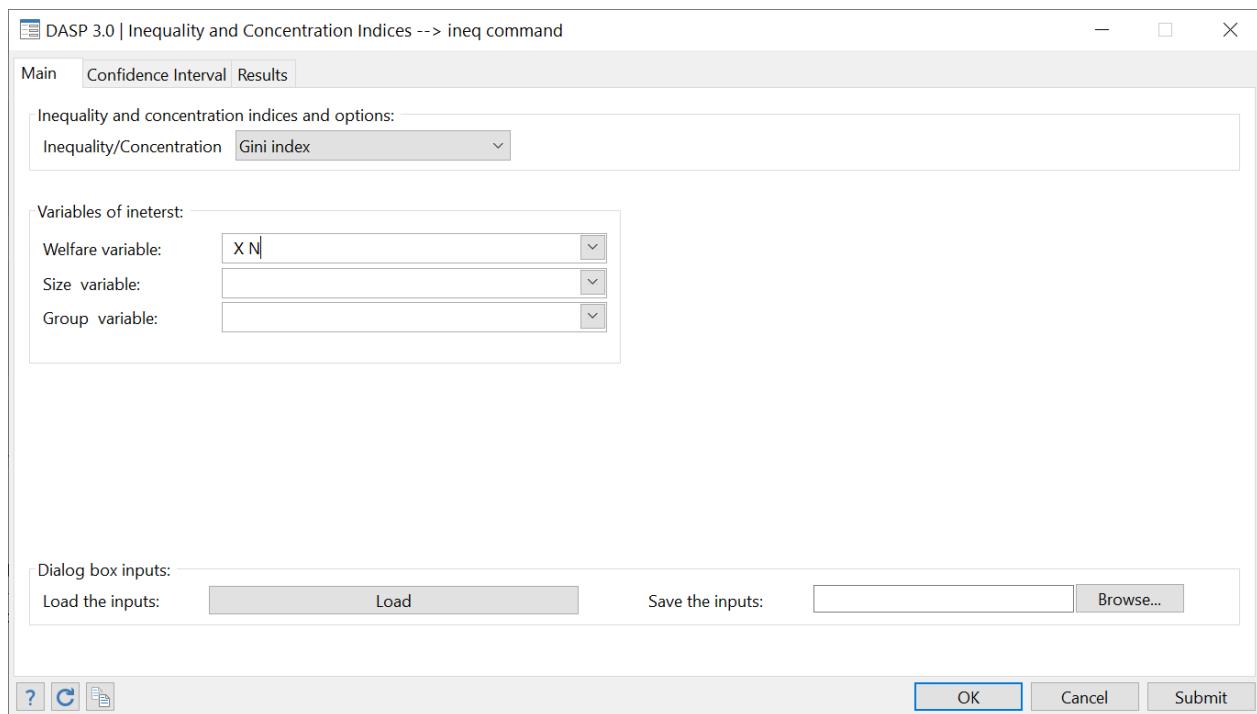
3. Estimate the difference in Burkina Faso's Gini index between 1998 and 1994
  - a. with variable of interest *expeqz* for 1998 and *expeq* for 1994;
  - b. with size variable set to *size*.

### Q.1

Steps:

- Type  
*sysuse can6.dta, clear*
- To open the relevant dialog box, type  
*db ineq*
- Choose variables and parameters as in

**Figure 54: Estimating Gini and concentration indices**



After clicking **SUBMIT**, the following results are obtained:

. ineq X N, index(gini)

Index : Gini index

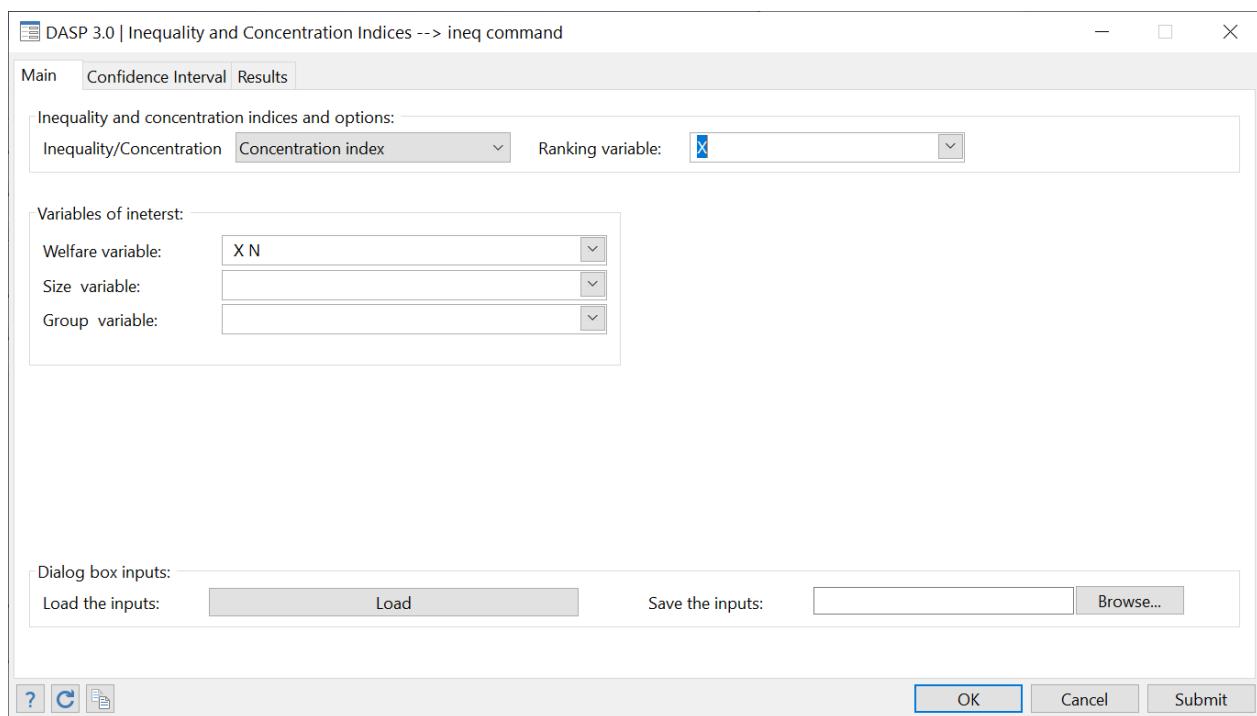
Variable	Estimate	STE	LB	UB
1: GINI_X	0.508456	0.016233	0.476602	0.540310
2: GINI_N	0.332355	0.012756	0.307323	0.357386

## Q.2

Steps:

- Choose variables and parameters as in the following dialog box:

**Figure 55: Estimating concentration indices**



After clicking **SUBMIT**, the following results are obtained:

```
. ineq X N, index(conc) rank(X)

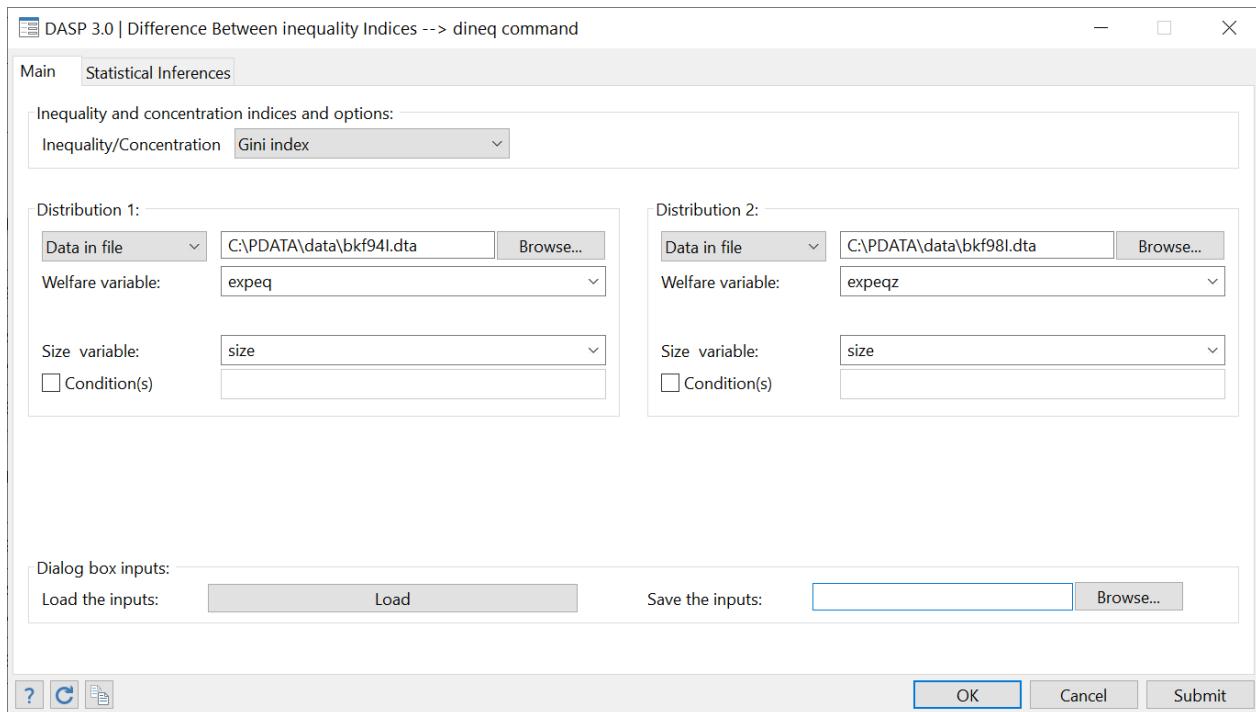
      Index          : Concentration index
      Ranking variable : X
      Variable   Estimate    STE     LB      UB
      1: GINI_X  0.508456  0.016233  0.476602  0.540310
      2: CONC_N  0.306050  0.013263  0.280024  0.332077
```

### Q.3

Steps:

- To open the relevant dialog box, type  
*db dineq*
- Choose variables and parameters as in the following dialog box:

**Figure 56: Estimating differences in Gini and concentration indices**



After clicking **SUBMIT**, the following information is obtained:

. dineq expeq expeqz, file1(C:\PDATA\data\bkf94I.dta) hsize1(size) file2(C:\PDATA\data\bkf98I.dta) hsize2(size) index(gini) Index : Gini index						
Index	Estimate	Std. Err.	t	P> t	[95% Conf. Interval]	
GINI_Dis1	.4500548	.0086095	52.2742	0.0000	.4331327	.4669769
GINI_Dis2	.4445629	.0128193	34.6792	0.0000	.419364	.4697618
diff.	-.0054919	.0154421	-.355645	0.7223	-.0358013 .0248175	

### 23.13 Using basic distributive tools

"What does the distribution of gross and net incomes look like in Canada?"

Using the **can6.dta** file,

1. Draw the density for gross income  $X$  and net income  $N$ .
  - The range for the x axis should be [0, 60 000].
2. Draw the quantile curves for gross income  $X$  and net income  $N$ .
  - The range of percentiles should be [0, 0.8]
3. Draw the expected tax/benefit according to gross income  $X$ .

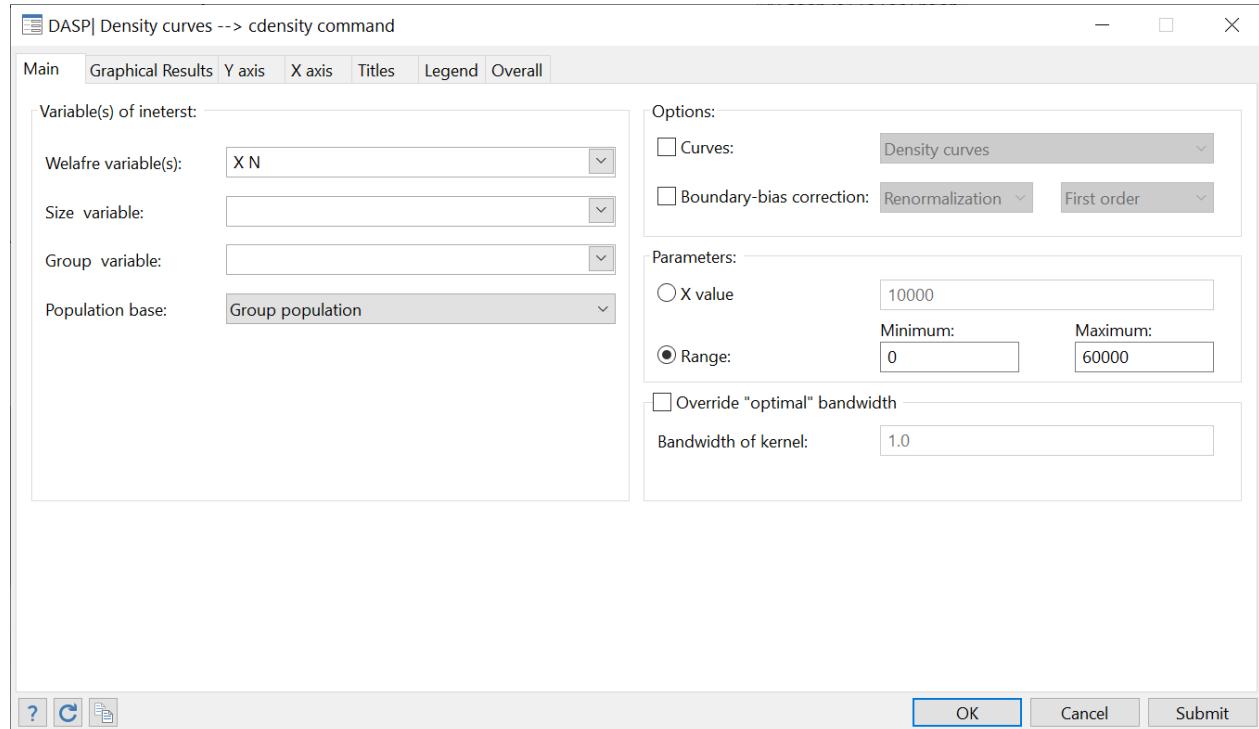
- The range for the x axis should be [0, 60 000]
  - Use a *local linear estimation* approach.
4. Estimate marginal rates for taxes and benefits according to gross income  $X$ .
- The range for the x axis should be [0, 60 000]
  - Use a *local linear estimation* approach.

### Q.1

Steps:

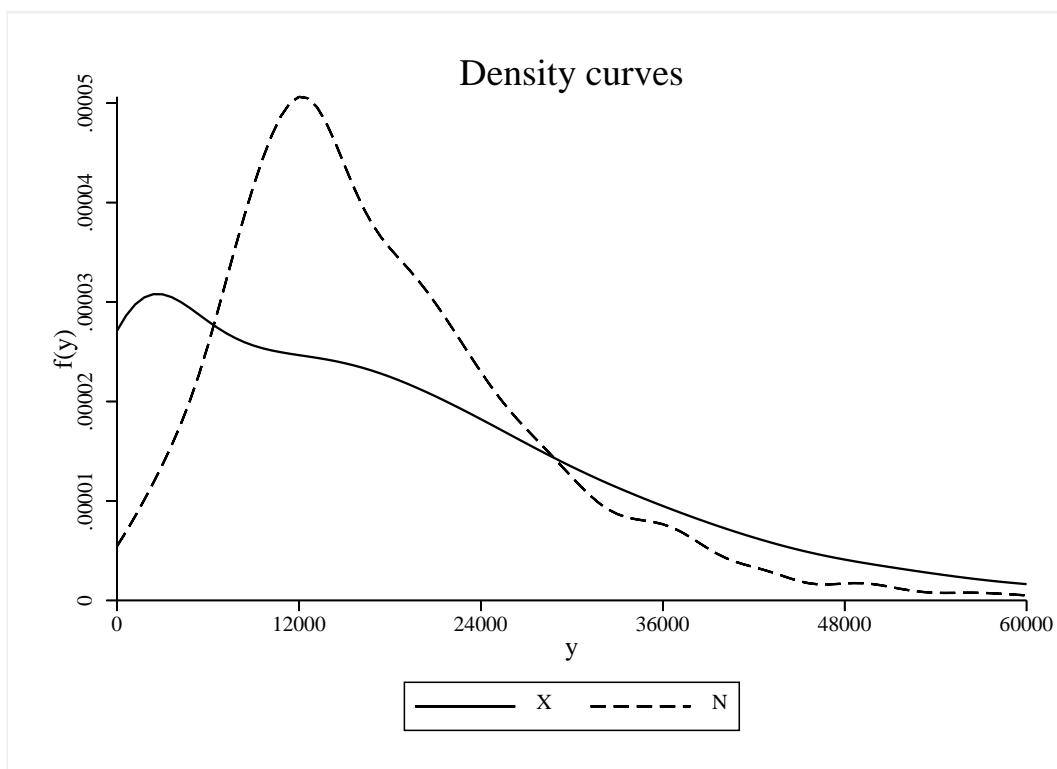
- Type  
`sysuse can6.dta, clear`
- To open the relevant dialog box, type  
`db cdensity`
- Choose variables and parameters as in the following dialog box:

**Figure 57: Drawing densities**



After clicking SUBMIT, the following appears:

Figure 58: Density curves

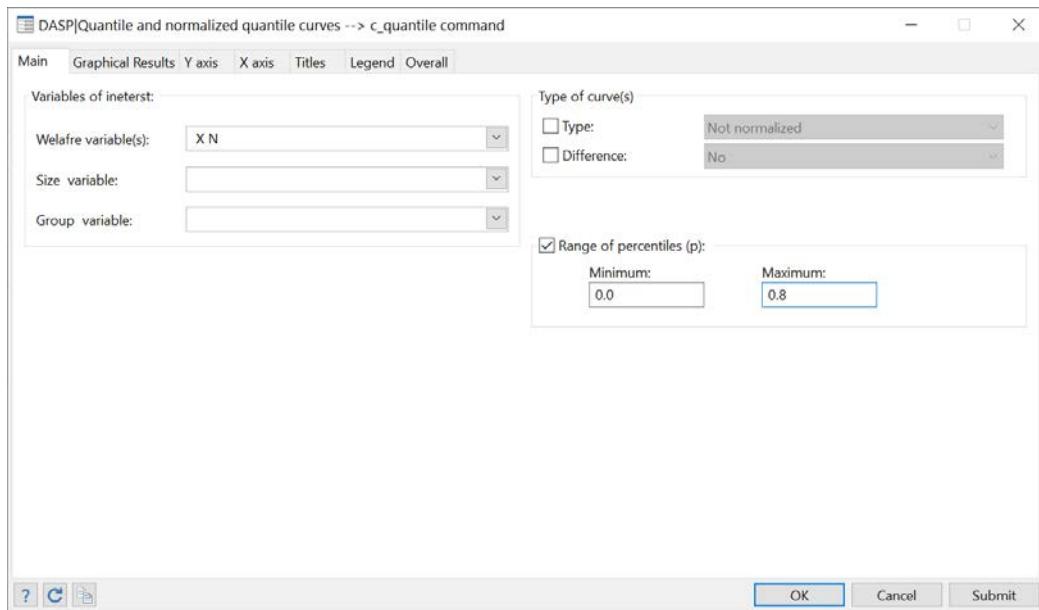


### Q.2

Steps:

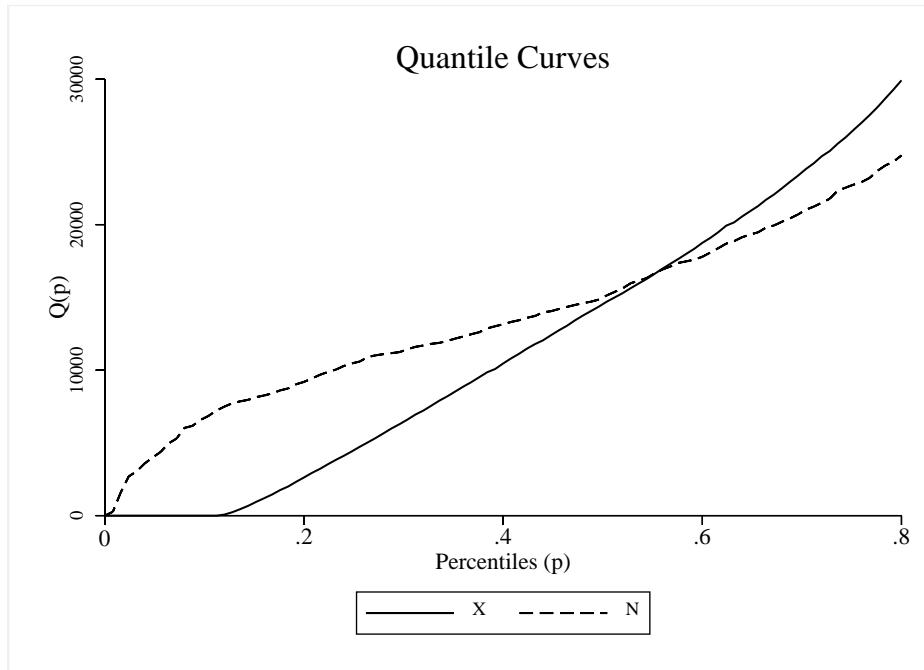
- To open the relevant dialog box, type : `db c_quantile`
- Choose variables and parameters as in the following dialog box:

**Figure 59: Drawing quantile curves**



After clicking SUBMIT, the following appears:

**Figure 60: Quantile curves**

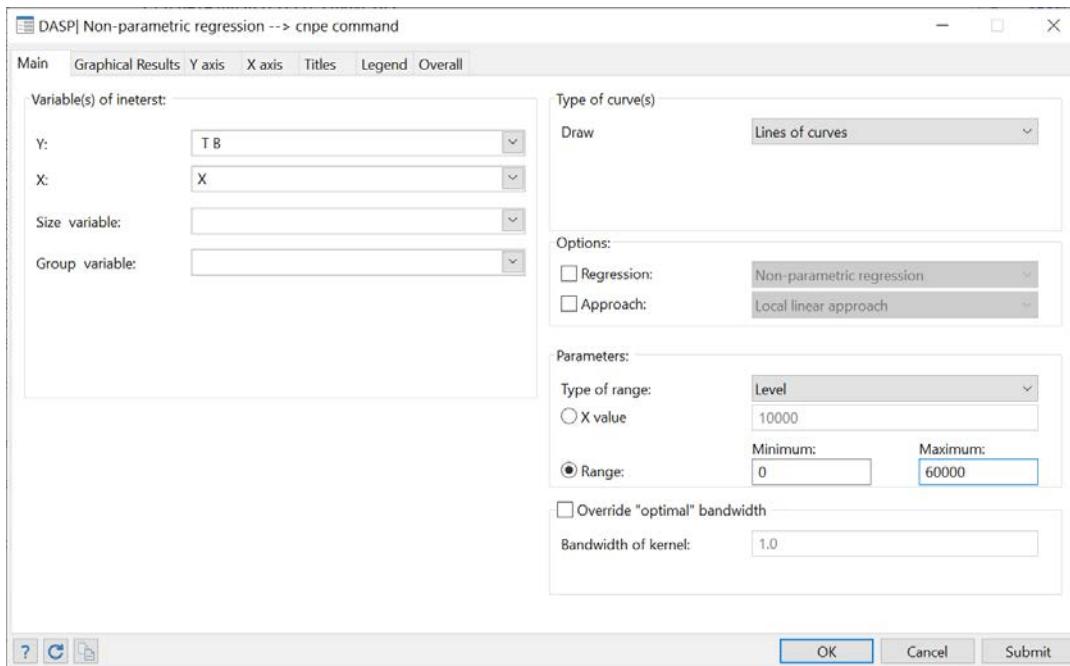


### Q.3

Steps:

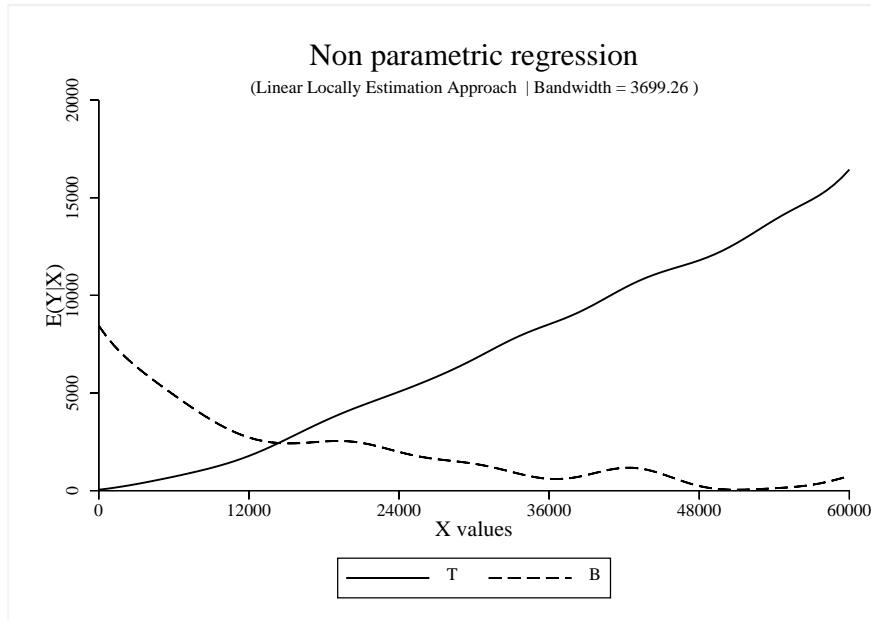
- To open the relevant dialog box, type : *db cnpe*
- Choose variables and parameters as in the following dialog box:

Figure 61: Drawing nonparametric regression curves



After clicking SUBMIT, the following appears:

Figure 62: Nonparametric regression curves

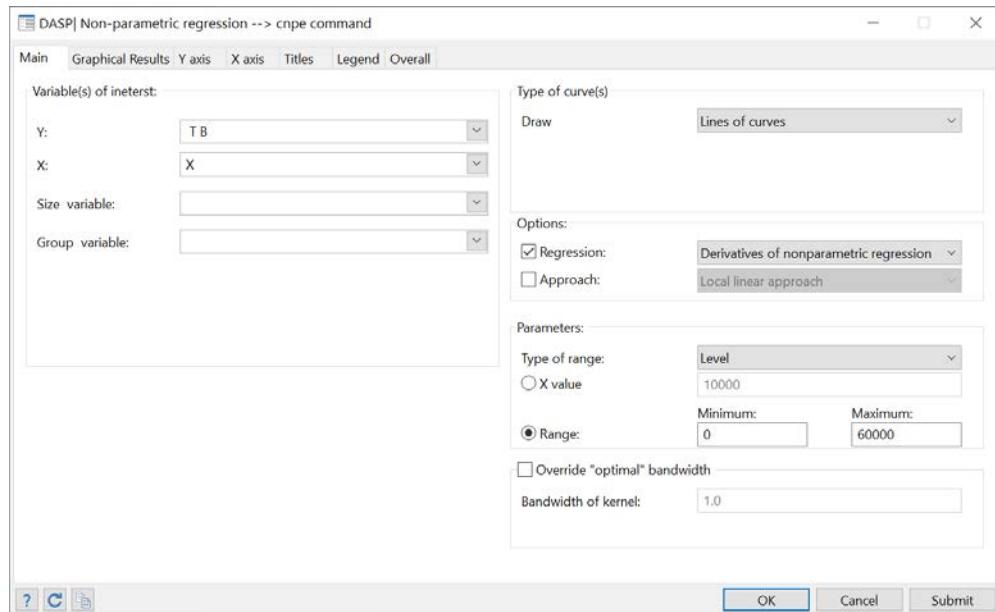


#### Q.4

Steps:

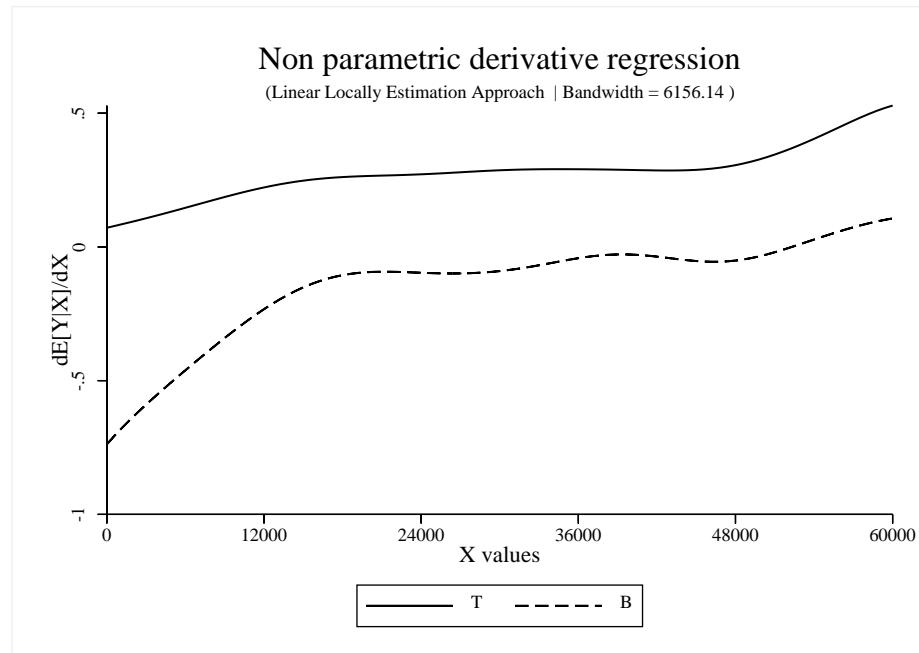
- Choose variables and parameters as in the following dialog box:

Figure 63: Drawing derivatives of nonparametric regression curves



After clicking SUBMIT, the following appears:

Figure 64: Derivatives of nonparametric regression curves



## 23.14 Plotting the joint density and joint distribution function

"What does the joint distribution of gross and net incomes look like in Canada?"

Using the **can6.dta** file,

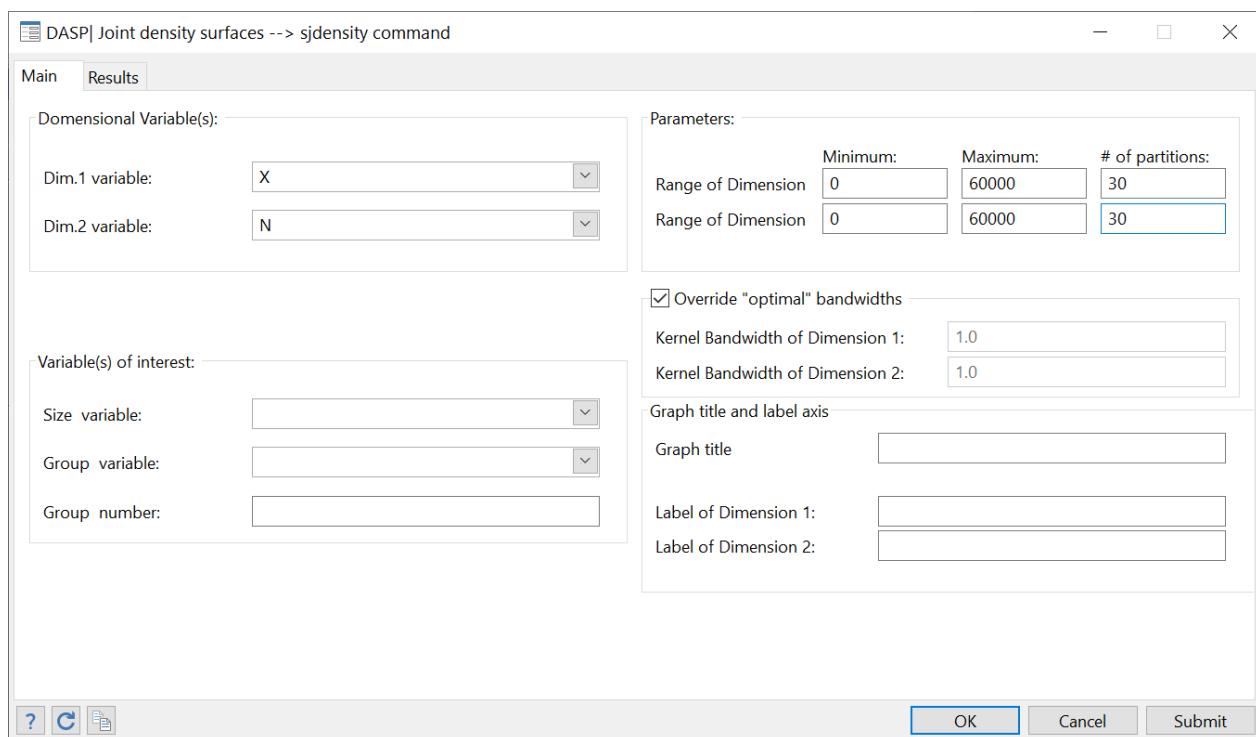
1. Estimate the joint density function for gross income  $X$  and net income  $N$ .
  - o X range : [0,60000]
  - o N range : [0,60000]
2. Estimate the joint distribution function for gross income  $X$  and net income  $N$ .
  - o X range : [0,60000]
  - o N range : [0,60000]

### Q.1

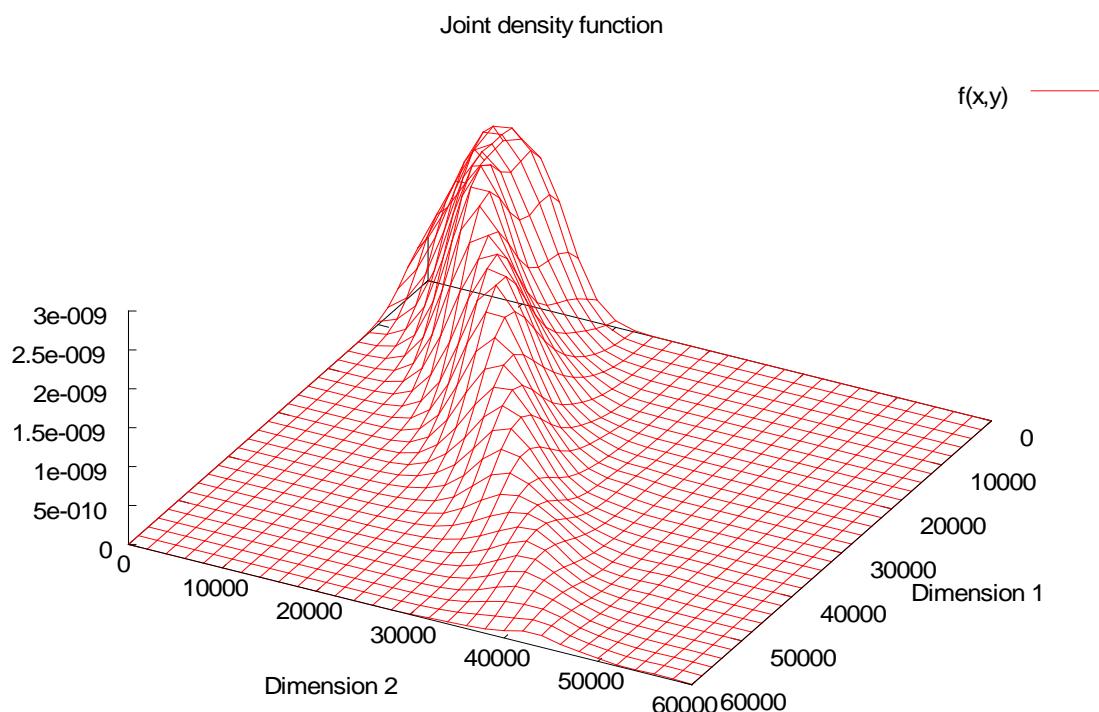
Steps:

- Type  
*sysuse can6.dta, clear*
- To open the relevant dialog box, type  
*db sjdensity*
- Choose variables and parameters as in

Figure 65: Plotting joint density function



After clicking SUBMIT, the following graph is plotted interactively with [Gnu Plot 4.2](#):

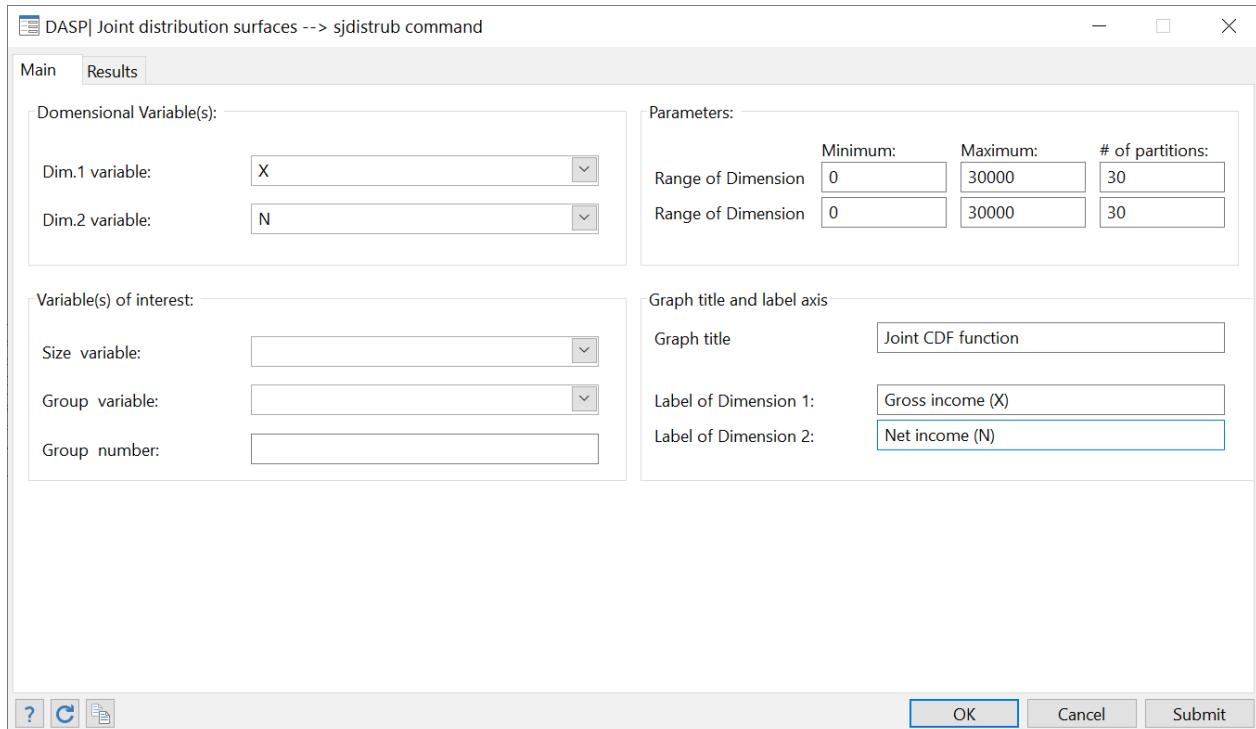


## Q.2

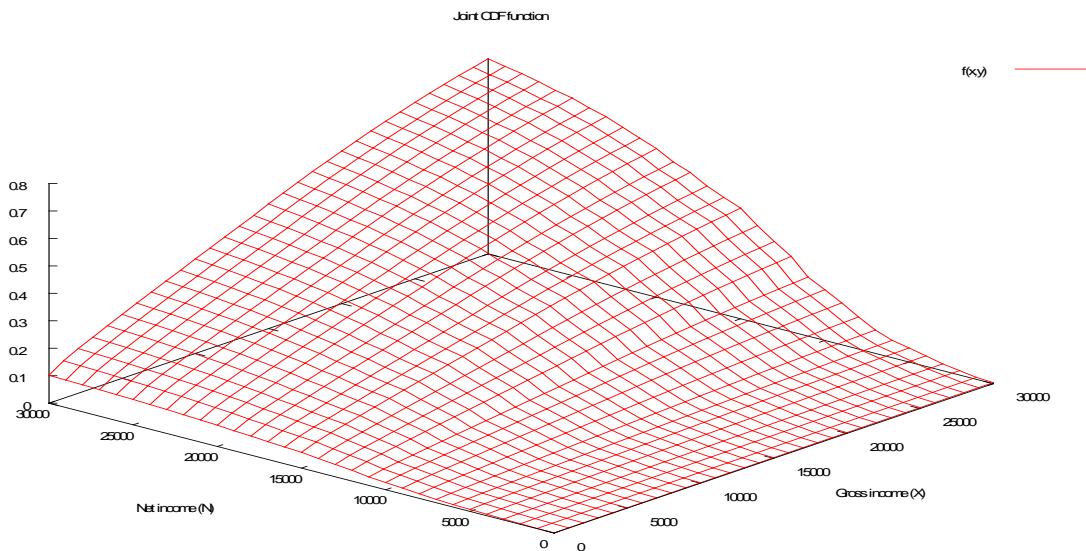
Steps:

- To open the relevant dialog box, type  
`db sjdistrub`
- Choose variables and parameters as in

Figure 66: Plotting joint distribution function



After clicking SUBMIT, the following graph is plotted interactively with [Gnu Plot 4.2](#):



### 23.15 Testing the bi-dimensional poverty dominance

Using the *columbia95l.dta* (distribution\_1) and the *dominican\_republic95l.dta* (distribution\_2) files,

1. Draw the difference between the bi-dimensional multiplicative FGT surfaces and the confidence interval of that difference when

	Var. of interest	Range	alpha_j
Dimension 1	haz : height-for-age	-3.0 / 6.0	0
Dimension 2	sprob : survival probability	0.8 / 1.0	0

2. Test for bi-dimensional poverty using the information above.

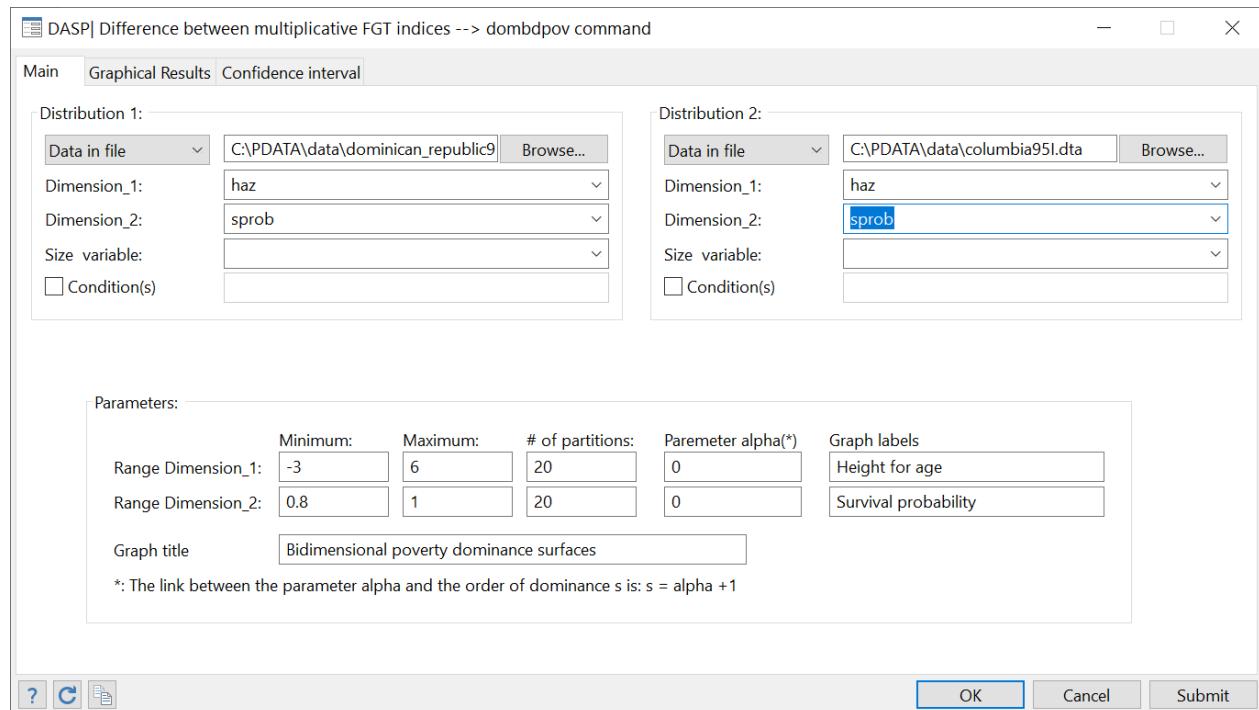
## Answers

### Q.1

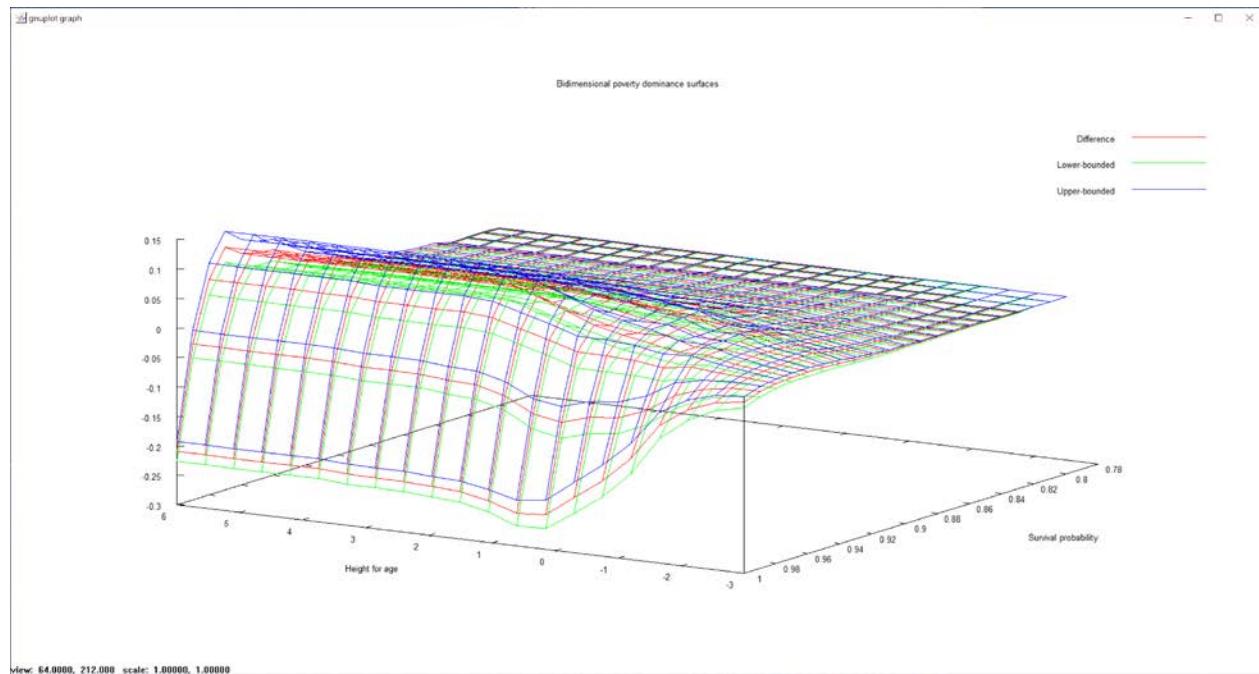
Steps:

- To open the relevant dialog box, type  
*db dombdpov*
- Choose variables and parameters as in

Figure 67: Testing for bi-dimensional poverty dominance



After clicking **SUBMIT**, the following graph is plotted interactively with [Gnu Plot 4.2](#):

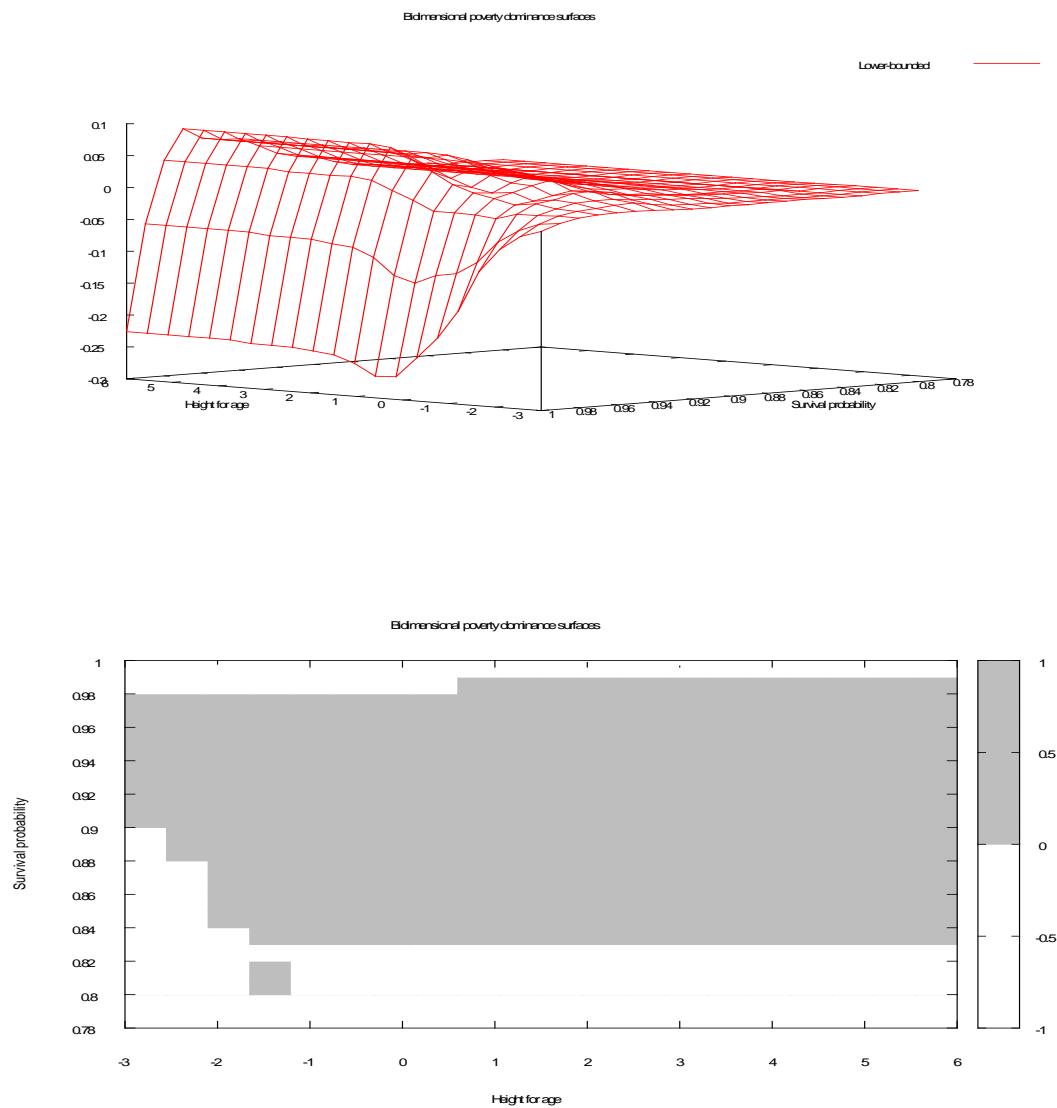


## Q.2

To make a simple test of multidimensional dominance, one should check if the lower-bounded confidence interval surface is always above zero for all combinations of relevant poverty lines – or conversely.

- For this, click on the panel “Confidence interval” and select the option lower bounded.
- Click again on the button Submit.

After clicking SUBMIT, the following graph is plotted interactively with [Gnu Plot 4.2](#):



## 23.16 Testing for pro-poorness of growth in Mexico

The three sub-samples used in these exercises are subsamples of 2000 observations drawn randomly from the three ENIGH Mexican household surveys for 1992, 1998 and 2004. Each of these three subsamples contains the following variables:

<i>strata</i>	The stratum
<i>psu</i>	The primary sampling unit
<i>weight</i>	Sampling weight
<i>inc</i>	Income
<i>hhsz</i>	Household size

1. Using the files *mex\_92\_2ml.dta* and *mex\_98\_2ml.dta*, and then draw the GIC curve.
2. Using the files *mex\_92\_2ml.dta* and *mex\_98\_2ml.dta*, test for first-order relative pro-poorness of growth when:
  - The primal approach is used.
  - The range of poverty lines is [0, 3000].
3. Repeat with the dual approach.
4. By using the files *mex\_98\_2ml.dta* and *mex\_04\_2ml.dta*, test for absolute second-order pro-poorness with the dual approach.
5. Using *mex\_98\_2ml.dta* and *mex\_04\_2ml.dta*, estimate the pro-poor indices of module *ipropoor*.
  - Parameter alpha set to 1.
  - Poverty line equal to 600.

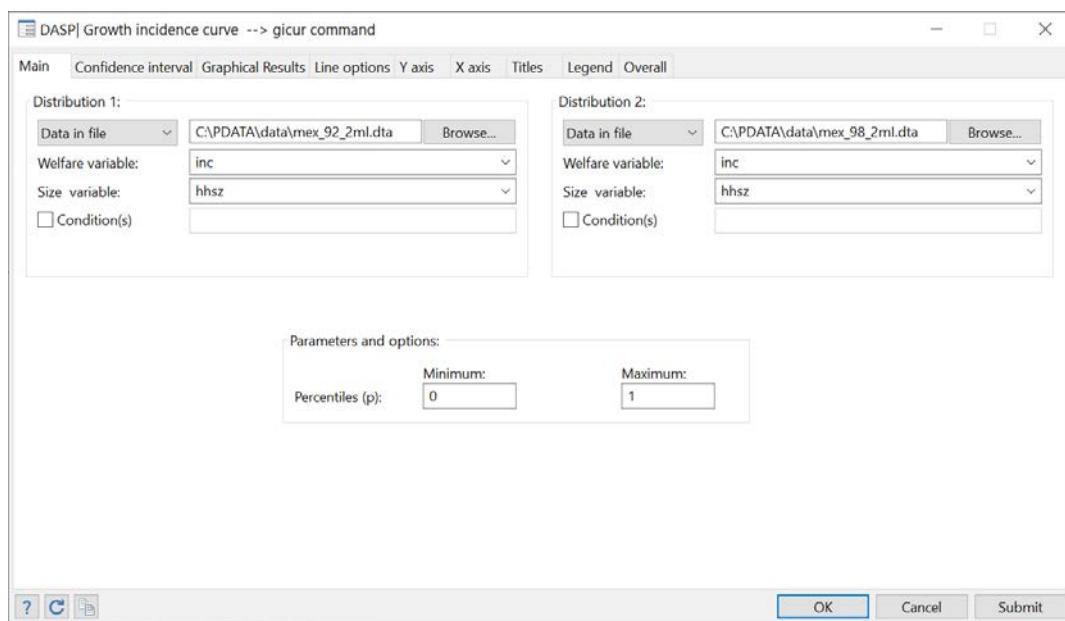
## Answers

### Q.1

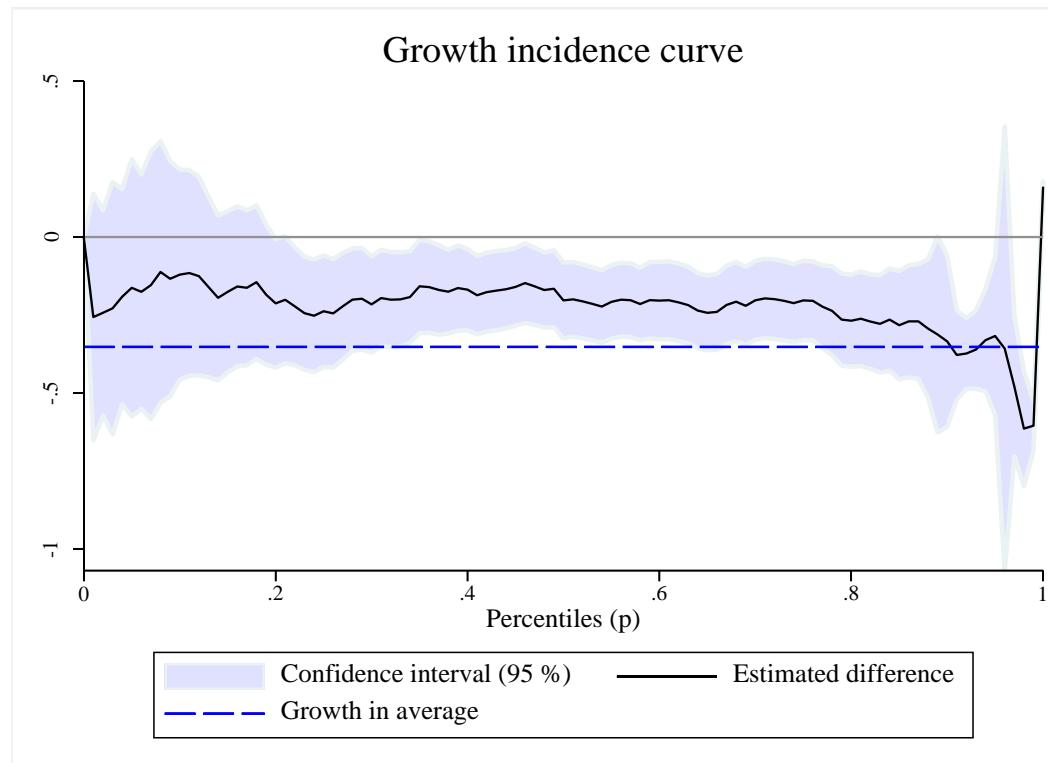
Steps:

- To open the relevant dialog box, type: *db gicur*
- Choose variables and parameters as in (select the upper-bounded option for the confidence interval):

**Figure 68: Drawing the growth incidence curve**



After clicking **SUBMIT**, the following graph appears

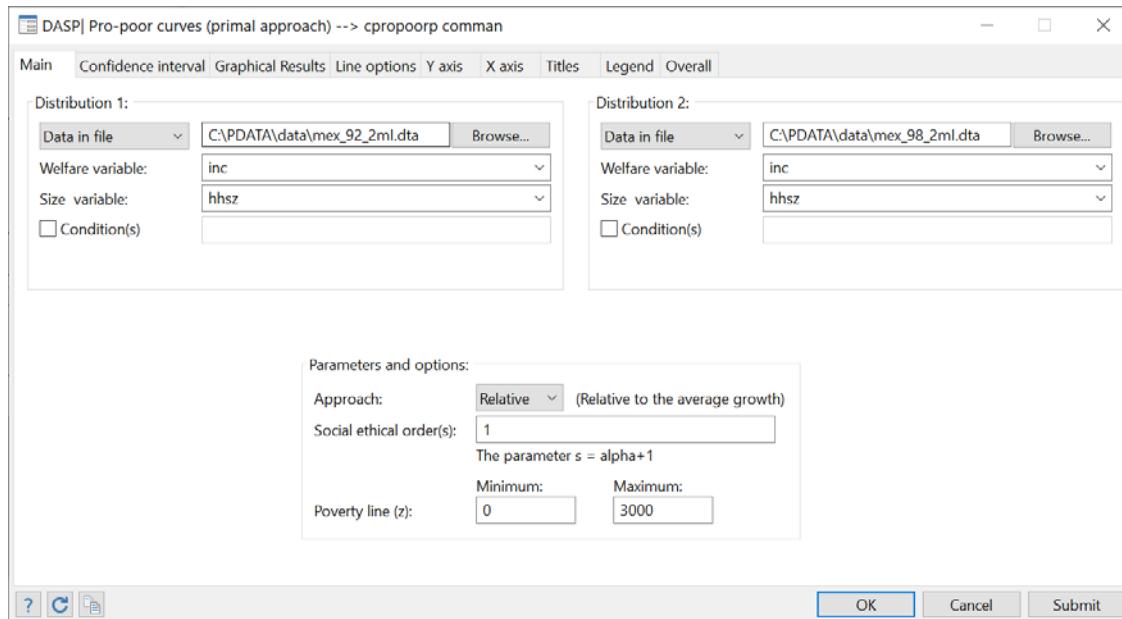


## Q.2

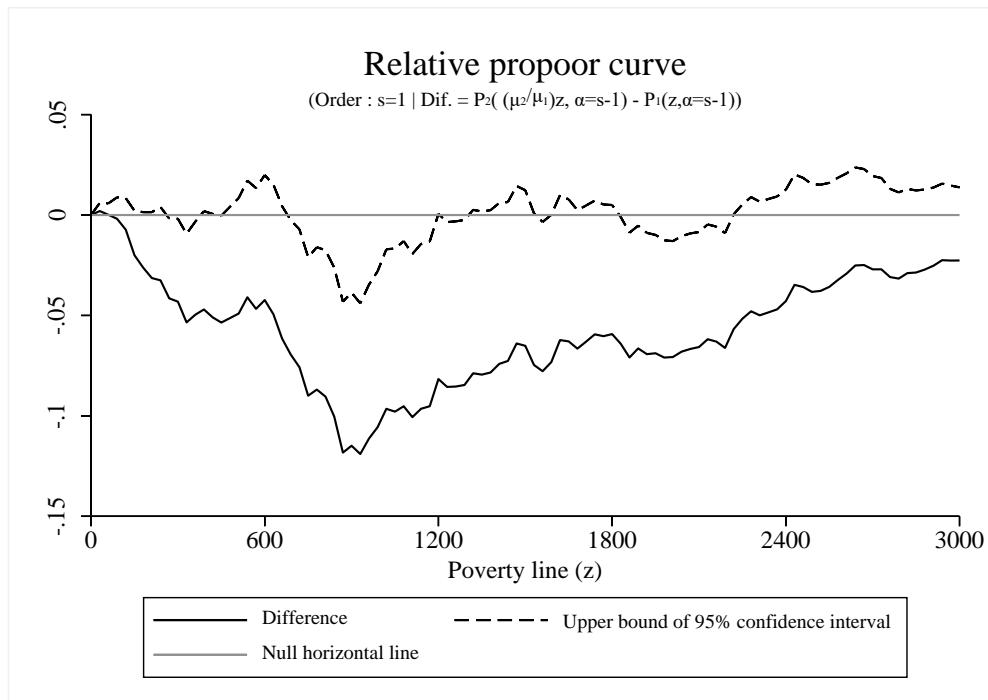
Steps:

- To open the relevant dialog box, type: `db cpropoorp`
- Choose variables and parameters as in (select the upper-bounded option for the confidence interval):

Figure 69: Testing the pro-poor growth (primal approach)



After clicking SUBMIT, the following graph appears

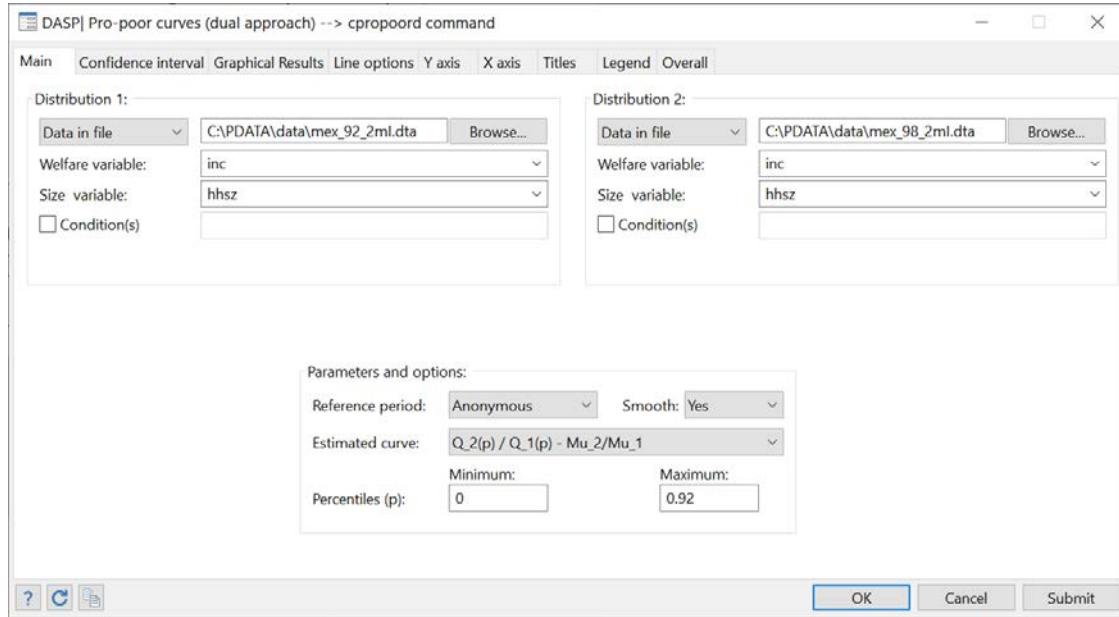


### Q.3

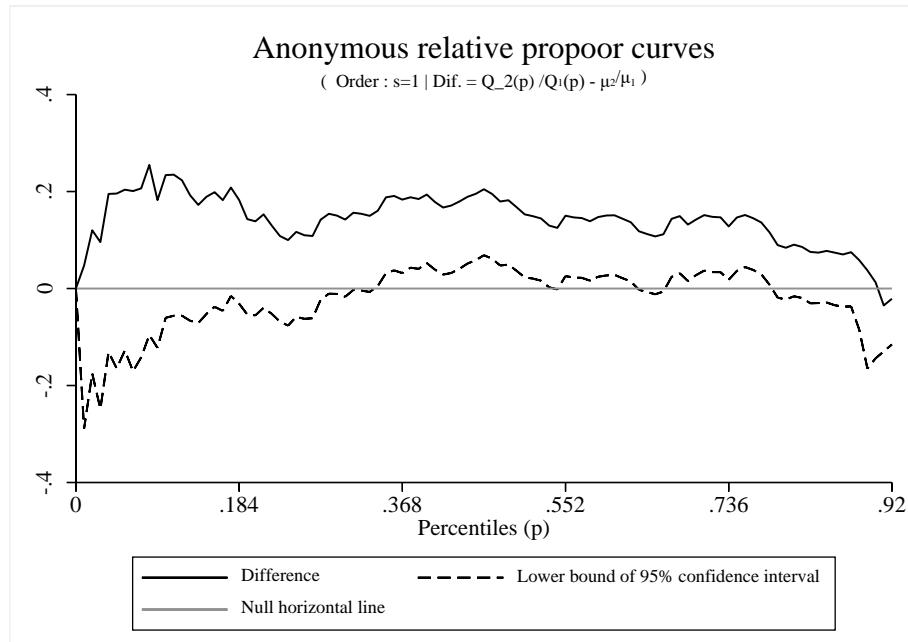
Steps:

- To open the relevant dialog box, type : db cpropoord
- Choose variables and parameters as in (with the lower-bounded option for the confidence interval):

Figure 70: Testing the pro-poor growth (dual approach)- A



After clicking SUBMIT, the following graph appears

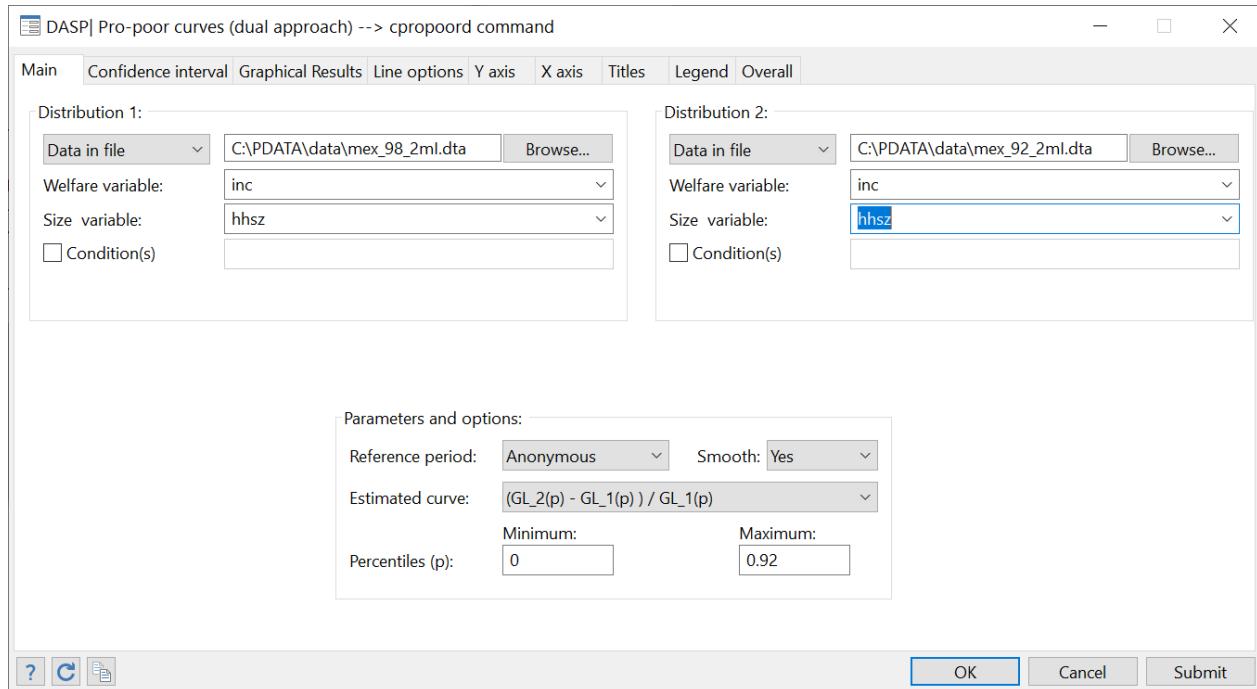


#### Q.4

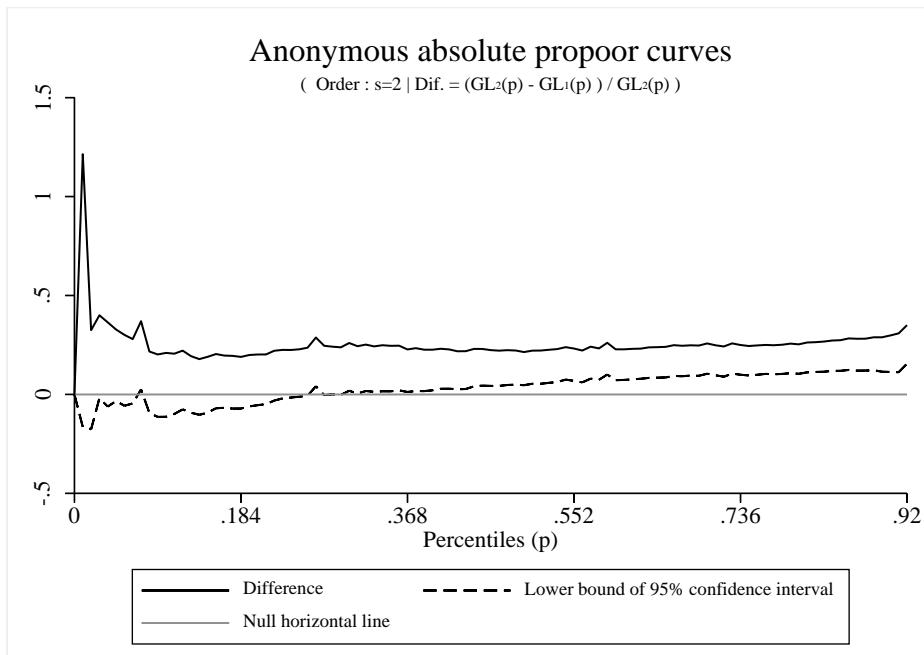
Steps:

- Choose variables and parameters as in (with the lower-bounded option for the confidence interval):

Figure 71: Testing the pro-poor growth (dual approach) – B



After clicking SUBMIT, the following graph appears

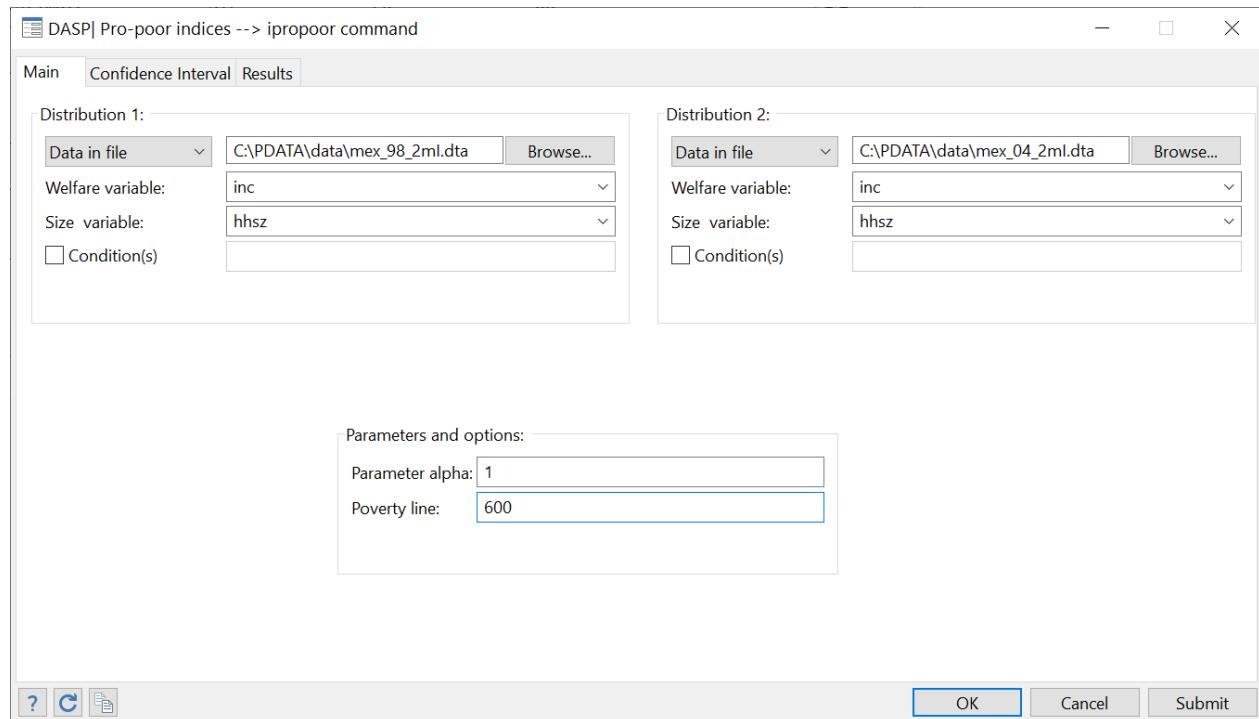


## Q.5

Steps:

- To open the relevant dialog box, type  
`db ipropoor`

- Choose variables and parameters as.



After clicking SUBMIT, the following results appear:

Poverty line : 600.00  
Parameter alpha : 1.00

Pro-poor indices	Estimate	STE	LB	UB
Growth rate(g)	0.582359	0.125512	0.336361	0.828357
Ravallion & Chen (2003) index	0.712285	0.052553	0.609283	0.815286
Ravallion & Chen (2003) - g	0.129926	0.128838	-0.122591	0.382443
Kakwani & Pernia (2000) index	1.325436	0.107047	1.115627	1.535244
PEGR index	0.771879	0.137331	0.502716	1.041042
PEGR - g	0.189520	0.049357	0.092783	0.286257

### 23.17 Benefit incidence analysis of public spending on education in Peru (1994)

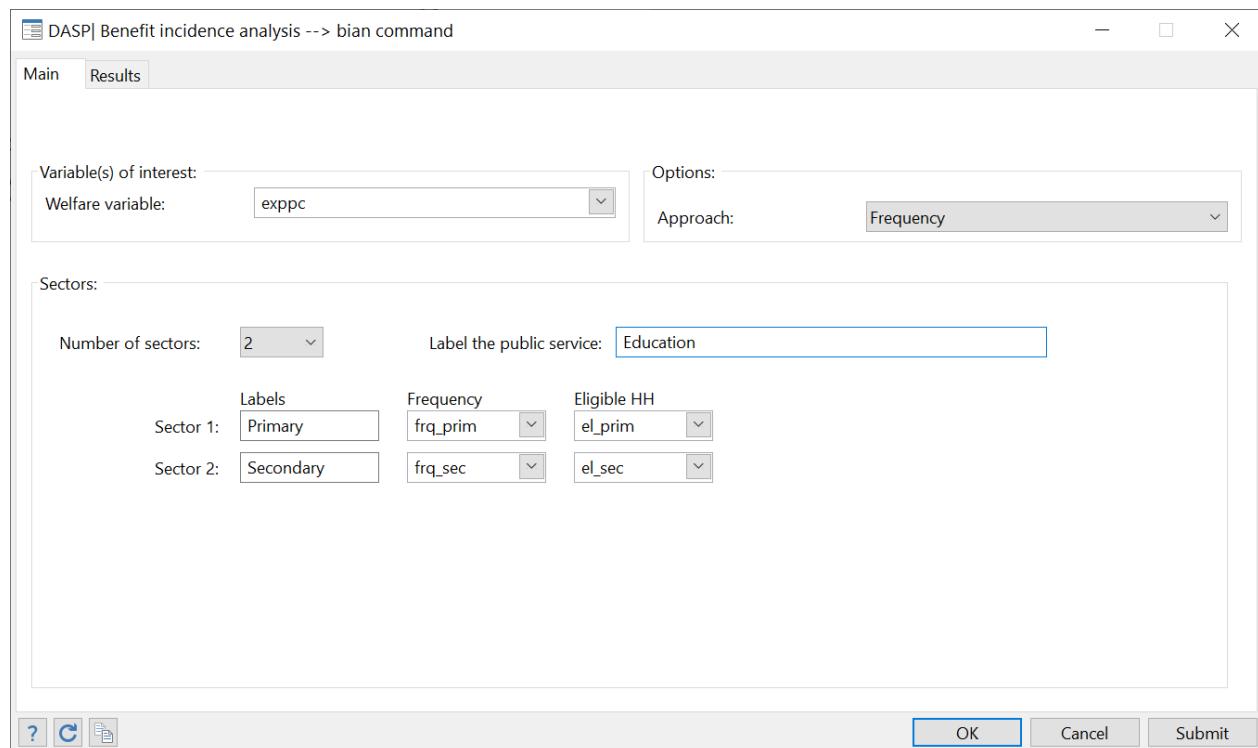
1. Using the peredu94l.dta file, estimate participation and coverage rates of two types of public spending on education when:
  - The standard of living is **exppc**
  - The number of household members that benefit from education is **fr\_prim** for the primary sector and **fr\_sec** for the secondary one.
  - The number of eligible household members is **el\_prim** for the primary sector and **el\_sec** for the secondary one.
  - Social groups are quintiles.

#### Answer

*sysuse perede94l.dta, replace*

Type **db bian** in the windows command and set variables and options as follows:

Figure 72: Benefit incidence analysis



After clicking on SUBMIT, the following appears:

```
. bian exppc, mainlab(Education) approach(1) frq1(frq_prim) sz1(el_pr)
_____
Benefit Incidence Analysis: Education
```

Share by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	0.218	0.226
Quintile 2	0.231	0.252
Quintile 3	0.221	0.225
Quintile 4	0.195	0.194
Quintile 5	0.135	0.102
All	1.000	1.000

Rate of Participation by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	0.762	0.325
Quintile 2	0.808	0.363
Quintile 3	0.772	0.323
Quintile 4	0.683	0.279
Quintile 5	0.472	0.147
All	0.699	0.287

2. To estimate total public expenditures on education by sector at the national level, the following macro information was used:

- Pre-primary and primary public education expenditure (as % of all levels), 1995: 35.2%
- Secondary public education expenditure (as % of all levels), 1995: 21.2%
- Tertiary public education expenditure (as % of all levels), 1995: 16%
- Public education expenditure (as % of GNP), 1995 = 3%
- GDP per capita: about 3 800.

Using this information, the following variables are generated

```
cap drop _var1;
gen _var1 = size*weight*3800;
qui sum _var1;
qui gen pri_pub_exp=0.03*0.352*r(sum)';
qui gen sec_pub_exp=0.03*0.212*r(sum)';
qui gen uni_pub_exp=0.03*0.160*r(sum)';
cap drop _var1;
```

- Total public expenditures on primary sector : **pri\_pub\_exp**
- Total public expenditures on secondary sector : **sec\_sec\_exp**
- Total public expenditures on university sector : **uni\_pub\_exp**

Estimate the average benefits per quintile and generate the benefit variables.

## Answer

Set variables and options as follows:

Figure 73: Benefit Incidence Analysis (unit cost approach)

DASP| Benefit incidence analysis --> bian command

Main Results

Variable(s) of interest:

Welfare variable: exppc Options:

Approach: Unit-cost benefit

Sectors:

Number of sectors: 2 Label the public service: Education

Sector 1:	Labels	Frequency	Eligible HH	Area indicator	Regional pub. expenditures
	Primary	frq_prim	el_prim		ri_pub_exp
Sector 2:	Secondary	frq_sec	el_sec		sc_pub_exp

?

OK Cancel Submit

DASP| Benefit incidence analysis --> bian command

Main Results

Result options:

Number of decimals: 3

Social groups Quintiles

Group variable:

Generate benefit variable(s):

Displayed results :

Share and rate of participation.

Average benefits.

Proportion of benefits.

?

OK Cancel Submit

After clicking on SUBMIT, the following appears:

---

Benefit Incidence Analysis: Education

---

Share by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	0.218	0.226
Quintile 2	0.231	0.252
Quintile 3	0.221	0.225
Quintile 4	0.195	0.194
Quintile 5	0.135	0.102
All	1.000	1.000

Rate of Participation by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	0.762	0.325
Quintile 2	0.808	0.363
Quintile 3	0.772	0.323
Quintile 4	0.683	0.279
Quintile 5	0.472	0.147
All	0.699	0.287

Average Benefits by Quintile Groups: (at the level of eligible members)

Groups	Primary	Secondary
Quintile 1	276.936	213.366
Quintile 2	293.796	238.235
Quintile 3	280.692	211.999
Quintile 4	248.121	183.199
Quintile 5	171.520	96.228
All	254.201	188.585

Average Benefits by Quintile Groups: (at the level of members that use the public service)

Groups	Primary	Secondary
Quintile 1	363.446	656.563
Quintile 2	363.446	656.563
Quintile 3	363.446	656.563
Quintile 4	363.446	656.563
Quintile 5	363.446	656.563
All	363.446	656.563

Proportion of Benefits by Quintile Groups and by Sectors.

Groups	Primary	Secondary
Quintile 1	0.136	0.085
Quintile 2	0.144	0.095
Quintile 3	0.138	0.085
Quintile 4	0.122	0.073
Quintile 5	0.084	0.038
All	0.624	0.376