Procedures for conducting poverty/distribution analysis of CGE simulation results with DAD.

John Cockburn, CIRPÉE, Université Laval, jcoc@ecn.ulaval.ca

20 octobre, 2005

1. Introduction

Freely available DAD software can be used in conjunction with Excel to analyse income distribution using standard money-metric techniques[1]. For an introduction, you may consult the following references: Ravallion, M. (1994), "Poverty Comparisons", Fundamentals of Pure and Applied Economics, Harwood Academic Publishers, Switzerland and Jean-Yves Duclos (2002), "Poverty and equity: Theory and Estimation", CREFA, Université Laval (we can email you upon request this last document). In particular, DAD is used to calculate standard distributional indicators and graph standard distributional curves based on vectors of household- or individual-level measures of income. We use DAD and Excel to compare income distribution in the base year and after various policy simulations. In this note, I outline the procedures I followed to carry out this distributional analysis in my microsimulation paper[2]. One exception is the proposed replacement of quantile curves and density functions by Lorenz curves.

2. Poverty analysis

In this section, we discuss[3]:

- Preparing data
- Calculation of FGT poverty indices: head count ratio, poverty gap, poverty severity (Table 9)
- Graphing FGT curves (Figures 1-3).
- Graphing Lorenz curves (not in microsimulation paper)
- Calculation of Atkinson and Gini inequality indices (Table 10).

2.1 Preparing data

In order to carry out distribution analysis, you will require household-level income or consumption data from a nationally-representative household survey. Household consumption is often preferered to household income for distribution analysis as it tends to be more stable and the data tend to be more reliable. DAD requires that these data be organized in an Ascii file containing the following information (separated by spaces, commas, semi-colons or colons)[4]:

- Household identification number
- Sample weights[5]

---

[1] The DAD software and an accompanying users' guide can be downloaded from the following web site: www.mimap.ecn.ulaval.ca. Online help is also available by choosing "?" "Help in HTML" within DAD.
[2] We can email this paper to you or you can download it: http://www.crefa.ecn.ulaval.ca/cahier/0118.pdf.
[3] Tables and figure numbers from the microsimulation paper are referenced in parentheses.
[4] This file can be prepared using Excel.
[5] Sample weights are generally provided at the household level for nationally-representative household

- A group variable (such as the household category) if you want to perform the analysis within sub-groups of households.
- Initial household income and/or consumption
- After-simulation household income and/or consumption

In the case of a CGE microsimulation model, the after-simulation income/consumption vectors are generated directly by the model. In the case of representative-household CGE models, you may generate the after-simulation household-level income and/or consumption values. To do so, first assign each household in your survey to one of the household categories used in your model. You can then increase the initial income/consumption values of each household by the average income/consumption variation of the corresponding household category as calculated by the CGE model simulation. This Ascii file can be read into DAD using the File Open command, choosing the file type "Ascii file (.txt, dat, prn)" and indicating the data separator used (comma, semi-colon or colon) in this file.

2.2 Calculation of FGT poverty indices

Foster-Greer-Thorbecke (FGT) indices are the most common poverty indices:

$$P_\alpha = \frac{1}{Nz^\alpha} \sum_{j=1}^{J} \left(z - y_j \right)^\alpha$$

where j is a sub-group of individuals with income below the poverty line (z), N is the total number of individuals in the sample, y_j is the income of individual j and α is a parameter that allows us to distinguish between the alternative FGT indices. When α is equal to 0, the expression simplifies to J/N or the headcount ratio, a measure of the incidence of poverty. Poverty depth is measured by the poverty gap, which is obtained with α equal to 1. The severity of poverty is measured by setting α equal to 2[6].

To calculate these indices with DAD, choose: "Poverty" and "FGT index". You can then choose to simultaneously calculate indicators for both the base year and any given simulation ("2 distributions" and check "independent distributions") or to calculate these indicators one at a time ("1 distribution"). Next, select the following information in the menu page that appears:

- "Variable of interest": Base-year and after-simulation household consumption or income
- "Weight variable": See footnote 5
- "Group variable" and "Group Number": If you want to do poverty analysis **within** household categories. In the micro-simulation paper, analysis was done by region as well as for the entire country.
- "Poverty line": You can compute this line ("compute line") based on the sample mean or quantile values if you want to do relative poverty analysis[7]. This is what was done in the microsimulation paper. However, it is preferable to use an absolute "basic needs" poverty line if this available (i.e. from your country's PMMA work).

---

surveys. To conduct an individual-level analysis instead of a household-level analysis, these sample weights should be multiplied by the number of individuals in each household. We propose the following standard equivalence scale (ES): $ES_i=1+0.7(Z_i-1-K_i)+0.5K_i$ where i is the household index, Z is the total number of household members and K is the number of children to calculate “adult-equivalent” income or expenditures.

[6] See Ravallion (1994) for a full discussion of poverty indicators.

[7] Note that the "Compute Line" window should be kept open when you choose this option.

- "Alpha" value: 0 (Headcount ratio), 1 (Poverty gap) or 2 (Poverty severity)
- "Standard deviation": We suggest that you do the analysis with standard deviations ("with STD")
- "Normalisation": We suggest that you choose "Normalised".

Once these options have been selected, click on "compute" to obtain the FGT indicator chosen ("Index Value") and the variation between the base year and the simulation results. You may then manually copy the results into a summary table such as Table 9 of the micro-simulation paper, before closing the results window (the "x" in the top right corner). This will bring you back to "FGT(Poverty)" window. This procedure must be repeated for each of the FGT indicators (alpha = 0,1,2) and, if you choose to perform the analysis within household categories, for each category as well as for all households taken together (i.e. without a "group variable").

2.3 Graphing FGT curves

As the FGT indicators are likely to be sensitive to the choice of poverty line, it is recommended to study these indicators for a whole range of poverty lines using the graph facility. Once again, choose: "Poverty" and "FGT index". The graphing facility is only available for "single distribution". The choice of menu options is identical as in the calculation of FGT indicators except that you have to study one income variable at a time and you should also select a range of poverty lines (e.g. from 0 to twice the "basic needs" poverty line). Click on "Range" and a range menu will appear at the bottom of the window. Once the range identified, click on "Graph" and, in the graph menu that appears, "Draw". You can then close the graph window.

This procedure can be performed first with the base year income values and then with the after-simulation income values, and the two curves compared. However the differences are often difficult to observe. For this reason, we prefer to study the **change** in FGT indicators directly. At the same time as DAD graphs the distribution of the chosen poverty indicator for a whole range of poverty lines, it also generates a series of 100 poverty line-indicator value pairs that is stored in memory (DAD stores these series for the last three graphs executed at any time). After graphing the FGT curves in the base year and after simulation, we can save these data pairs to an Excel file. To do this, close the graph window if it is still open, select the "graph" option at the top-right of the "FGT(Poverty)" window, click on "Reset Graph", then "Save XY" and give a name to the file in question (e.g. "data.xls").

Without closing DAD, this file can then be opened in Excel (as "read-only", "delimited", "Next", "Space", "Finish") and the data can be copied into template Excel files we have prepared to automatically graph changes in the FGT curves[8]: FGT0_template.xls, FGT1_template.xls and FGT2_template.xls (see attached; we suggest that you keep backup versions of these files). Once the data have been successfully copied to the template file, this file should be saved with a new name. The same process can be followed for the other FGT indicators and, when applicable, the other household sub-groups.

[8] The data should be copied into the range A2:D102 of the template files.

2.4 Graphing Lorenz curves

Whereas the FGT indices are used for poverty analysis, the Lorenz curve is an indicator of inequality. With households in rising order of income, it expresses the cumulative percentage of population on the x-axis (the p-values) and the cumulative percentage of income received on the y-axis.

The procedure is basically the same as for FGT curves. To graph Lorenz curves, choose "Curves" "Lorenz" "1 distribution" in the main DAD menu and select the following options in the Lorenz Curve menu:

- "Variable of interest": Base-year and after-simulation household consumption or income
- "Weight variable": See footnote 5
- "Group variable and "Group Number": If you want to do poverty analysis **within** household categories.
- "P-value": It is not necessary to choose a p-value as the graph shows all p-values
- "Standard deviation": We suggest that you do the analysis with standard deviations ("with STD")
- "Range": By default, the full-range of p-values (0 to 1) will be analysed so there is no need to define a range.

After completing this analysis for the base year and after-simulation values, close the graph window if it is still open, select the "graph" option at the top-right of the "Lorenz Curve" window, click on "Save XY" and give a name to the file in question (e.g. "data.xls"). This file can then be opened in Excel and copied to the Lorenz_template.xls file (see attached) following the same procedure used for the FGT curves. Once the data have been successfully copied to the template file, this file should be saved with a new name. The same process can be followed, when applicable, for the other household sub-groups and for all households taken together.

2.5 Calculation of Gini and Atkinson inequality indices

The Gini coefficient provides a compact expression of the Lorenz curve. It is equal to the area between the Lorenz curve and the 45 degree line divided by the total area under the 45 degree line. The Gini coefficient can be calculated by selecting "Inequality", "S-Gini Index". S-Gini indices are a generalization of the Gini indices; the usual Gini index is obtained as a particular case by setting rho=2, "2 distributions" and "independent distributions" in the main DAD menu and completing the following options in the S-Gini (Inequality) menu:

- "Variable of interest": Base-year and after-simulation household consumption or income
- "Weight variable": See footnote 5
- "Group variable" and "Group Number": If you want to do poverty analysis **within** household categories.
- "rho": set to 2 (the default setting)
- "Standard deviation": We suggest that you do the analysis with standard deviations ("with STD")

You may then manually copy the results into a summary table such as Table 10 of the micro-simulation paper, before closing the results window (the "x" in the top right corner). This will bring you back to "S-Gini (Inequality)" window. If you choose to perform the analysis within household categories, this procedure must be repeated for each one as well as for all

households taken together (i.e. without a "group variable").

The Atkinson inequality index is an alternative to the Gini coefficient. It is equal to:

$1-(y_e/u)$

where y_e is the uniform income level that, if received by all households, would generate the same total welfare as the actual income distribution, and u is the actual mean income level:

$$y_e = \left(\sum_{i=1}^{n} f(y_i) y_i^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}$$

It can be calculated by selecting "Inequality", "Atkinson Index", "2 distributions" and "independent distributions" in the main DAD menu and completing the following options in the Atkinson (Inequality) menu:

- "Variable of interest": Base-year and after-simulation household consumption or income
- "Weight variable": See footnote 5
- "Group variable" and "Group Number": If you want to do poverty analysis **within** household categories.
- "epsilon": This is the societal aversion to inequality parameter. Higher epsilon-values are associated with higher aversion. Typical values range from 0.25 to 2.
- "Standard deviation": We suggest that you do the analysis with standard deviations ("with STD")

You may then manually copy the results into a summary table such as Table 10 of the micro-simulation paper, before closing the results window (the "x" in the top right corner). This will bring you back to "Atkinson (Inequality)" window. If you choose to perform the analysis within household categories, this procedure must be repeated for each one as well as for all households taken together (i.e. without a "group variable")