



Software review

DAD, an innovative tool for income distribution analysis

QI ZHANG

The University of Chicago, 5841 S. Maryland Ave. MC 2007, Chicago, IL 60637, USA

Numerous theoretical and empirical studies have examined the economics of income distribution and redistribution. Kolm [5] and Atkinson's [1] original papers initiated extensive theoretical discussions on techniques to measure income inequality. Lambert's [6] milestone book completely reviewed the theoretical basis underlying income distribution and related issues including redistribution, social welfare, and poverty. With the development of theory in income distribution, economists are eager to apply the theory to empirical problems in welfare economics, labor economics, and development economics [2]. Income related data become the key component of all empirical analysis. How to analyze the income data correctly and efficiently is the challenge to every researcher in this field. Hardware seems no longer to be a problem after the information technology advance from the outdated IBM mainframes to personal computers that are much faster and smaller nowadays. However, there is no single dominant software for income distribution analysis. Popular statistical packages, like SAS, Stata, and GAUSS, were not developed specifically for economists who are interested in income inequality, redistribution and social welfare. These software packages do not provide the most effective way to analyze income data for the following reasons:

- (1) Users must transform complex mathematical formulas into programs to calculate special indices or statistics. It is an arduous task to program or debug pages of code, especially for an economist who is more interested in the results. Although the Windows technology is widely used in developing those statistical packages, the interface is still not adequately user-friendly, and the learning process often proves to be long and inefficient.
- (2) The use of different packages by different researchers can hinder the exchange of ideas and lower research efficiency. Because there are software-specific rules of coding, it is time consuming, sometimes even impossible, to transform the developed program from one computer platform to another. It is also possible that those developed codes are not public domain so that researchers have to code their own programs from scratch. Clearly, it is inefficient to reinvent the

wheel. Therefore, it is beneficial for researchers to have a tool tailored for income distribution analysis. DAD 4.2 (Distributive Analysis/Analyse Distributive), a software package developed by Duclos, Araar, and Fortin, is currently the most appropriate software for use in this field.

In brief, DAD has handy menus embedded to execute the following major functions:

- (1) Calculating indices frequently used by researchers, including inequality indices, poverty indices, social welfare indices, and redistribution indices;
- (2) Graphing various curves related to those indices, such as Lorenz curves and concentration curves;
- (3) Analyzing those curves by applying stochastic dominance theory;
- (4) Decomposing inequality, poverty, and progressivity across population subgroups or by structural components;
- (5) Providing asymptotic standard deviations on all estimates of distributive analysis.

DAD covers most regular computation and graphing of inequality, poverty, and social welfare discussed in the literature. Without developing complicated codes, economists can obtain the results they are interested in with a few clicks and put more emphasis on the interpretation of results. The software allows researchers to be more productive.

Compared with the traditional statistical software, DAD has the following advantages:

- (1) Accounting for complex sampling design. Household surveys rarely use simple random sampling technique, which draws the sample observations directly and randomly from the population with a certain sampling weight. Instead, household surveys often use a complex sampling that draws the sample observations through stratification and clustering. As pointed out by Deaton [3], ignoring complex sampling design could result in biased estimates of population statistics. However, with the exception of Stata, very few statistical packages take the complex sampling design into consideration. It is potentially risky to use those packages as they may generate biased estimates. DAD has a unique sampling design feature listed under "Edit" menu. Users can conveniently set the attributes of the sampling design. Moreover, in every application, users can choose between simple random sampling (SRS) and complex sampling design
- (2) Allowing two distributions to be processed simultaneously. Before executing any application, users can choose to apply one or two distributions. If two distributions are selected, users can specify whether the two distributions are independent or not to calculate the standard errors. Users can easily compare the outputs from two distributions.
- (3) Including methodology of distributive analysis in the Help menu. DAD's Help menu serves as a quick reference for economists interested in mathematical

formula or theory. It is an extra benefit for users to have a built-in resource to consult for instruction in income distribution analysis.

DAD4.2 is not the final version of the software. A few minor changes are suggested to improve the quality of the software:

- (1) Some basic editing features should be included in future versions of the software. Due to the wide usage of Excel, users are accustomed to basic editing tools, such as Undo, Cut, Find, and Select All Cells in one column/row. Although DAD has one function of computing column, it is far less powerful than Excel in generating new columns/rows from old columns/rows. Users may not be comfortable editing cells/columns/rows after data is imported into DAD. Though it is unrealistic to include every feature available in Excel, users expect the most common functions.
- (2) DAD has strict requirements on the completeness of the imported data. One missing value may ruin a calculation. The software has a feature in “Import Data Wizard” to replace all missing values with a certain number (text not allowed) or drop the entire row/column. But sometimes users are interested in locating those missing values first and then deciding how to deal with them. One option is to replace all missing values with a weird number, say -9999 . However, because no “Find” function exists under the “Edit” menu, it is painstaking to find a specific cell in a large table. My suggestion is to allow text in the replacement value in the import wizard so that users can set a marker, like “No Value”, to search.
- (3) More work needs to be done on the graphs. In DAD4.2, the default labels of x and y axes are “Value X” and “Value Y”. These labels make it difficult for users to understand the auto-generated curves. Moreover, the software uses “curve # x ” to refer to the different curves plotted, which is also confusing. Since the type of curves is fixed in DAD, the developer could have more specific default labels for two axes and refer to curves in a more understandable way.
- (4) An index of key words in the user’s menu and theoretical menu could be helpful for users. To enhance the ease of use of the software, the developers can create a basic interactive learning tool for new users.
- (5) Although standard deviations are computed for all estimates, no formal inference testing is provided. The software will be more powerful if an independent function of “Inference Testing” is added into the tool bar.

Although DAD4.2 still has a few limitations, it is a very efficient tool to conduct income distribution analysis. Graduate students, junior researchers, and even experienced researchers who do not like coding benefit most from using this innovative software in their research. The developers have made a great contribution to the empirical analysis of income distribution. It is a long journey for the developers to refine their software as more new estimation or statistical techniques continue to emerge from the literature.

References

1. Atkinson, A. B.: On the measurement of inequality, *J. Econom. Theory* **2** (1970), 244–263.
2. Atkinson, A. B.: Bringing income distribution in from the cold, *The Economic J.* **107**(441) (1997), 297–321.
3. Deaton, A.: *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, Johns Hopkins University Press, Baltimore, MD, 1997.
4. Duclos, J. Y., Araar, A. and Fortin, C.: DAD: A software for distributive analysis/analyse distributive, MIMAP programme, International Development Research Centre, Government of Canada, and CIRPÉE, Université Laval (www.mimap.ecn.ulaval.ca), 2003.
5. Kolm, S. C.: The optimal production of social justice, In: J. Margolis and H. Guitton (eds), *Public Economics*, Macmillan, London, 1969.
6. Lambert, P. J.: *The Distribution and Redistribution of Income: A Mathematical Analysis*, Manchester University Press, Manchester, U.K., 1993.