

September 2014

## Response to the comments made by Professor Stephen Jenkins

UNU-WIDER is very grateful to Professor Stephen Jenkins of the London School of Economics<sup>1</sup> for his most useful comments and suggestions for improvement regarding the World Income Inequality Database (WIID), version WIID2.oc.

We have now updated the WIID and also revised the data, attempting to do our best to take into account all the useful comments by Professor Jenkins.

Below are listed suggestions made by Prof Jenkins which have been taken care of in the revised version of WIID (WIID3.ob). The table below details these specific comments and actions taken:

Issues Identified by Jenkins	Actions
Country documentation missing for the UK and Vietnam. (Page 8)	Country documentation provided.
Data not sorted by country year. (Page 9, 1st para)	Now sorted.
Currency reference should be broken into 2 variables. Currency unit and reference period. (Page 9)	Completed.
Misspellings in variables should be corrected including typographical errors. (Page 9)	Has been corrected.
Reported Gini shows 2 redundant decimal points. (Page 10, 1st para)	Has been removed.
All numeric variables need to be labeled. E.g., AreaCovr, PopCovr, AgeCovr, IncSharU, UofAnala, Equivsc, IncDefn. All these should be numeric with values labeled. (Page 10, 1st para)	All the mentioned variables have been numerically coded. The IncDefn is now renamed WelfareDefn.
Year variable is not numeric but string as there are observations such as 1953-55. In these cases, mid points should be taken and mentioned in the source comments. (Page 10, 1st para)	This has been corrected in the revised version.
AK variable has missing values. (Page 10, 1st para)	Variable has been deleted.
Two-letter country identifiers should be provided. (Page 10, 2nd para)	Done.
Country identifier should also be numeric. This will mean that we give the country codes some unique numbers. (Page 10, 2nd para)	The country codes are now numeric and values have been given country ISO code names. E.g., Afghanistan code is 1 and the label is AFG.
The quality variable needs to be value labeled. E.g., 1= best quality and 4=worst. (Page 10/11)	Quality variable is now defined for all observations in the data. 4=not known and not worst, however.
Country documentation of all countries should be in one single pdf file. At present each country's documentation has to be downloaded separately from a drop down menu. (Page 10, 2nd para)	Country documents are now available both separately and as a single file.

<sup>1</sup> Jenkins, S.P. (2014). 'World Income Inequality Databases: An assessment of WIID and SWIID'. Working Paper 2014-31. Colchester: Institute for Social and Economic Research, University of Essex. <https://www.iser.essex.ac.uk/publications/working-papers/iser/2014-31>. Also forthcoming in F. Ferreira and N. Lustig (Eds) (2014). *Journal of Economic Inequality*, special issue on 'Appraising Cross-National Income Inequality Databases'.

Country documentation and variable documentation need to be closely integrated. Any changes made to variables should be mentioned both in variable documentation and country documentation. (Page 11, 2nd para)	This is the type of work that adds value to the WIID and we try to deal with this as well as we can. It has been reconciled to a large extent and will be an ongoing process.
Remove the Gini variable (computed) as there is more than 90 per cent correlation with the ReportedGini variable from the dataset. (Page 12, 2nd para)	This has been dropped in the revised version. Only ReportedGini is included in the new version and it is renamed simply as Gini.
In around 70 per cent of cells, there are at least two observations per country-year, and about one fifth have five or more observations. The prevalence of multiple observations is greatest in the 1990s. (Page 13, 1st para)	We have dropped some multiple observations after checking the original source. Others are strictly not duplicates, that is, sometimes the source is different and area coverage is different.
Jäntti series for Finland gives different values. (Page 13, 2nd para)	This series has been dropped.
It may be appropriate to incorporate Uusitalo observation into a Finnish series, especially since there is no 1985 estimate from the 'Statistics Finland 2005' series. (Page 13/14. Sec 3.3, para 1)	The Uusitalo series used OECD scale whereas Statistics Finland series use modified OECD scale so we will not merge the two for the time being.
There are some country estimates in WIID, which are out of line compared to LIS and EUSTAT estimates. E.g., Hungary for 1990s etc. (Page 15, 1st para)	In WIID we give all sources of information available, but we can indicate the harmonized as the main, preferred observations (reflected in quality ratings).
Overlapping information and that which adds no new information should not be appended to the WIID (as earlier suggested by Atkinson and Brandolini). (Page 15)	Has been done in the case of LIS, SEDLAC, EUSTAT. When there was overlap of information, only the latest information was appended.
In case of USA, CPS has changed its methodology over the years. There are differences in trends due to different definition of income used. Jenkins has highlighted this point with regard having multiple observations for a particular year. Jenkins points out that the cost is that the income definition is not commonly used in cross-national comparative analysis nowadays. (Page 16, Section 3.4)	We keep multiple series, but indicate the preferred one in quality ratings / remarks in the data set to guide the user to adopt the best practice. A note is mentioned in the country documentation about this point.
In case of EU countries, EUSTAT Gini's should be preferred to WIID Gini's as the former is harmonized and the latter is not. (Page 37/38)	We report both, and details of EUSTAT and OECD methodology in the user guide section.