World Income Inequality Database Version 3.3
User Guide and Data Sources
Contents

Preface .......................................................................................................................... 3

The basic principles behind WIID3 ............................................................................. 4
  The conceptual base ................................................................................................. 4
  Income or consumption? ......................................................................................... 4
  The income concept ................................................................................................. 5
  The consumption/expenditure concept .................................................................. 5
  Other conceptual issues ......................................................................................... 7
  Information regarding OECD and EUROstat databases ....................................... 7

The construction of WIID3 ......................................................................................... 9
  New observations .................................................................................................... 9
  Corrected observations ......................................................................................... 10
  Format of the database ......................................................................................... 11

The documentation .................................................................................................. 11
  The documentation in the database itself ............................................................ 11
  The country information sheets ............................................................................ 13

The quality rating ..................................................................................................... 13
  The criteria used .................................................................................................. 14
  The final rating ..................................................................................................... 15

Some final guidelines ............................................................................................... 16

List of Variables ...................................................................................................... 17

Glossary .................................................................................................................... 18

References ................................................................................................................. 24
Preface

In the UNU-WIDER World Income Inequality Database—known by its acronym WIID—information on income inequality for developed, developing, and transition countries is stored.

WIID was initially compiled in 1997-99 for the UNU-WIDER-UNDP project ‘Rising Income Inequality and Poverty Reduction: Are They Compatible?’ directed by Professor Giovanni Andrea Cornia, the then Director of UNU-WIDER. This resulted in WIID version 1.0, published in September 2000. The database was subsequently updated as part of the UNU-WIDER project ‘Global Trends in Inequality and Poverty’ led by Professor Tony Shorrocks, who was the UNU-WIDER Director from 2001-09. This update was called WIID2, and its latest revision was released in 2008.

The current revision—WIID3.3—is an annual update of the third major revision of the database, WIID3. It is part of the 2014-18 UNU-WIDER programme of work on ‘Transformation, Inclusion, and Sustainability’. The current version retains the basic structure of WIID2, but corrects for a number of inconsistencies and other issues found in the earlier version. Most importantly, the current update includes observations for eight more years, with the latest observations now reaching the year 2013.

The new dataset was prepared by a WIDER team including Tony Addison, Gyanendra Badgaiyan, Nina Badgaiyan, Miguel Niño-Zarazúa, Milla Nyyssölä, Jukka Pirttilä, and Finn Tarp. During the process, useful comments were received from Professor Markus Jäntti (Stockholm University), Professor Stephen Jenkins (London School of Economics), and Tony Shorrocks (Global Economic Perspectives Ltd), for which we are grateful. Professor Jenkins provided a thorough review of WIID2 (Jenkins 2015). We have tried our best to take into account all his helpful comments regarding how WIID should be developed. Badgaiyan et al. (2015) describes our response in more detail.

Please refer to the data set as:


Finn Tarp
Director, UNU-WIDER
Helsinki
30 September 2015
The basic principles behind WIID3

The conceptual base

There are no easy to use income/consumption distribution data. Unlike national accounts data which are in principle comparable across countries, there is no agreed basis of definition for the construction of distribution data. Sources and methods might vary, especially across but also within countries. This may be the case even if the data comes from the same source. In their influential articles on the use of secondary data in studies of income distribution, Atkinson & Brandolini (2001, 2009) discuss quality and consistency in income distribution data both within and across countries. They show how both levels and trends in distributional data can be affected by data choices. In light of this, it is not an easy task to construct a secondary database with distribution data. To get some structure, we started by defining a preferred set of features for the conceptual base and the underlying data. With the conceptual base we mean the definitions of income or consumption/expenditure, the statistical units to be adopted, the use of equivalence scales and weighting.

Income or consumption?

The first issue to address is whether inequality estimates based on income or consumption should be preferred. According to Deaton & Zaidi (2002) the empirical literature on the relationship between income and consumption has established, for both rich and poor countries, that consumption is not closely tied to short-term fluctuations in income, and that consumption is smoother and less variable than income. Especially in developing countries, where the rural agriculture sector is large, it is difficult to gather accurate income data. Accordingly, consumption data should be used. Atkinson & Bourguignon (2000) do not share this view. There is, according to them, no clear advantage in using consumption rather than income in studying distributional issues. The use of consumption rather than income data raises problems of definition and observation, the main conceptual problem being the treatment of durables and the necessity of imputing value for their services.

Regardless of the different views, the collection of inequality observations is restricted to what in practice is available. In most industrialized countries inequality and poverty are assessed with reference to income, not consumption (Deaton & Zaid, 2002). This tradition is followed in much of Latin America. By contrast, most Asian and African surveys have always collected detailed consumption data. The fact that distribution data can be based on both income and consumption is the first step in the construction of comparable statistics. In WIID2 we have strived to collect observations with reference to both income and consumption, whenever it is possible.
The income concept

The second issue is how to define income and consumption. As stated earlier, there is no agreed basis of definition as in the case of national accounts data. Concerning income data, some steps have been taken towards developing international standards. The Final Report and Recommendations of the Canberra Group (2001) provides an appropriate base for defining the most preferred income concept as the objective of the group was to enhance national household income statistics by developing standards on conceptual and practical issues related to the production of income distribution statistics. Even if the work of the group is mainly based on OECD-country experience, we believe that the main conclusions concerning the income concept also hold for other countries. In Table 1, the income concept as recommended by the Canberra Group for international comparisons of income distribution is given. The definition of total and disposable income as recommended by the group should include certain components to be considered complete. We have been drawing special attention to whether the underlying income concept includes income items such as imputed rents for owner-occupied dwellings\textsuperscript{1}, imputed incomes from home production and in-kind income in general. Imputed rent from owner-occupied dwellings is not mentioned in the concept of the Canberra group since many countries do not provide estimates for this item, and it is differently valued in different countries. Imputed rents should, however, preferable be included even if the comparability between countries might suffer somewhat. Home production and in-kind income are crucial in developing and transition countries. The income concept cannot be considered complete for these countries if income in-kind and income from home production are not included. The inequality indices reported will in the first place be those calculated on the basis of disposable income, but if indices based on earnings or gross incomes (total income according to the Canberra Group terminology) are available, they will also be reported.

The consumption/expenditure concept

On the consumption side, the situation is more difficult. Deaton & Zaidi (2002) from the LSMS-group at the World Bank\textsuperscript{2} have worked out some guidelines. Their recommendations on how to use consumption data for welfare measurement were used. Where the Canberra Group recommendations were built mainly on OECD-country experience, these recommendations are mainly built on experiences from developing countries. The crucial thing here is to evaluate the consumption rather than to simply calculate the expenditures. In other words to make a distinction between what is consumed and what is purchased. This means that one is not interested in the purchase value of durable goods but in the use or rental value. As is clear from Table 1, taxes paid, purchase of assets, repayments of loans and lumpy expenditures should not be included in the consumption aggregate. If they are included, we refer to expenditure rather than consumption. Again we have paid attention to the inclusion of non-monetary items.

---

\textsuperscript{1} Please refer to the glossary for an explanation of the terms used.
\textsuperscript{2} LSMS stands for Living Standards Measurement Study. The household surveys provided by this study can be found at http://econ.worldbank.org/WEBSITE/EXTERNAL/EXTDEC/EXTRSEACHER/EXTLSMS/0,,menuPK:3359053-pagePK:64168427-piPK:64168435-theSitePK:3358997,00.html
<table>
<thead>
<tr>
<th><strong>Table 1 Preferred set of underlying concepts for inequality estimates in WIID2</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The income concept</strong> recommended by the Canberra Group for international comparisons of income distribution:</td>
</tr>
<tr>
<td>1. Employee income</td>
</tr>
<tr>
<td>Cash wages and salaries</td>
</tr>
<tr>
<td>2. Income from self-employment</td>
</tr>
<tr>
<td>Profit/loss from unincorporated enterprise</td>
</tr>
<tr>
<td>Imputed income from self-employment</td>
</tr>
<tr>
<td>Goods and services produced for barter, less cost of inputs</td>
</tr>
<tr>
<td>Goods produce for home consumption, less cost of inputs</td>
</tr>
<tr>
<td>3. Income less expenses from rentals, except rent of land</td>
</tr>
<tr>
<td>4. Property Income</td>
</tr>
<tr>
<td>Interest received less interest paid</td>
</tr>
<tr>
<td>Dividends</td>
</tr>
<tr>
<td>5. Current transfers received</td>
</tr>
<tr>
<td>Social insurance benefits from employers’ schemes</td>
</tr>
<tr>
<td>Social insurance benefits in cash from government schemes</td>
</tr>
<tr>
<td>Universal social assistance benefits in cash from government</td>
</tr>
<tr>
<td>Mean-tested social assistance benefits in cash from government</td>
</tr>
<tr>
<td>Regular inter-household cash transfers received</td>
</tr>
<tr>
<td>6. Total income (sum of 1 to 5)</td>
</tr>
<tr>
<td>7. Current transfers paid</td>
</tr>
<tr>
<td>Employees’ social contributions</td>
</tr>
<tr>
<td>Taxes on income</td>
</tr>
<tr>
<td>8. Disposable income (6 less 7)</td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>The consumption aggregate</strong> recommended by Deaton &amp; Zaidi (2002) for welfare measurements:</td>
</tr>
<tr>
<td>1. Food consumption</td>
</tr>
<tr>
<td>Food purchased from market</td>
</tr>
<tr>
<td>Home produced</td>
</tr>
<tr>
<td>Received as gift or in kind payment</td>
</tr>
<tr>
<td>2. Non-food consumption</td>
</tr>
<tr>
<td>Daily use items</td>
</tr>
<tr>
<td>Clothing and house wares</td>
</tr>
<tr>
<td>Health expenses</td>
</tr>
<tr>
<td>Education expenses</td>
</tr>
<tr>
<td>Transport</td>
</tr>
<tr>
<td>3. Durable goods</td>
</tr>
<tr>
<td>The use-value (rental value) of durables</td>
</tr>
<tr>
<td>4. Housing</td>
</tr>
<tr>
<td>Rents paid</td>
</tr>
<tr>
<td>If dwelling is owned by household or received free of charge, an estimate of the rental equivalent (imputed rent)</td>
</tr>
<tr>
<td>Utilities (water, electricity, garbage collection etc.)</td>
</tr>
<tr>
<td><strong>To be excluded:</strong> Taxes paid, purchase of assets, repayments of loans and lumpy expenditures. If durables are included with their purchase value or/taxes paid, purchase of assets, repayments of loans and lumpy expenditures, the concept to be referred to is <strong>expenditures.</strong></td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Other conceptual issues:</strong></td>
</tr>
<tr>
<td>1. Household should be the basic statistical unit</td>
</tr>
<tr>
<td>2. Per capita incomes or consumption/expenditure should be measured</td>
</tr>
<tr>
<td>3. Person weights should be applied</td>
</tr>
</tbody>
</table>
Other conceptual issues

The third issue to look at concerns other conceptual issues. Here we follow quite closely the recommendations of the Canberra Group. Departures from the recommendations are mainly driven by practical matters.

a) The household should be the basic statistical unit; the statistical unit for analysis of economic well-being has to be one where assumptions of sharing of economic resources are most plausible. The Canberra Group motivates the preference for the household by the relationship of households to both micro (survey) and macro (SNA) data uses. In practice, households are often used as the basic statistical unit. The different definitions of households that appear in the data are a problem which will affect the estimates and users should be aware of.

b) Income or consumption should be adjusted to take account of household size, using per capita incomes or consumption. The Canberra Group suggests the use of equivalence scales as the relative need of different sized households is different. We decided to choose per capita estimates as the preferred ones, as they are the one mostly commonly available and since a lot of different equivalence scales are in use which weakens the comparability of the estimates.

c) Person weights are preferred as the users of income statistics most often are concerned with the economic well-being of individuals and not with the well-being of households.

Estimates not following the preferred set of definitions are not automatically considered to be of bad quality, but when updates were made, the definitions were followed whenever we could make a choice. Due to unavailability of observations using the preferred set of definitions, estimates based on other definitions were in several cases used. The differences appear especially in the statistical units and in the weighting.

Information regarding OECD and EUROstat databases

WIID collects many observations from the OECD and EUROstat. OECD Database:

The OECD Income Distribution database (IDD)\(^3\) has been developed to benchmark and monitor countries’ performance in the field of income inequality and poverty. It contains a number of standardised indicators based on the central concept of “equivalised household disposable income”, i.e. the total income received by the households.

households less the current taxes and transfers they pay, adjusted for household size with an equivalence scale. While household income is only one of the factors shaping people’s economic well-being, it is also the one for which comparable data for all OECD countries are most common. Income distribution has a long-standing tradition among household-level statistics, with regular data collections going back to the 1980s (and sometimes earlier) in many OECD countries.

Achieving comparability in this field is a challenge, as national practices differ widely in terms of concepts, measures, and statistical sources. In order to maximise international comparability as well as inter-temporal consistency of data, the IDD data collection and compilation process is based on a common set of statistical conventions (e.g. on income concepts and components). The information obtained by the OECD through a network of national data providers, via a standardized questionnaire, is based on national sources that are deemed to be most representative for each country.

The EU-Statistics on Income and Living Conditions (EU-SILC) instrument is the EU reference source for comparative statistics on income distribution and social inclusion at the European level. It provides two types of annual data for 27 European Union countries, Croatia, Iceland, Norway, Switzerland and Turkey:

- Cross-sectional data pertaining to a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions, and
- Longitudinal data pertaining to individual-level changes over time, observed periodically over a four year period.

EU-SILC does not rely on a common questionnaire or a survey but on the idea of a “framework”. The latter defines the harmonised lists of target primary (annual) and secondary (every four years or less frequently) variables to be transmitted to Eurostat; common guidelines and procedures; common concepts (household and income) and classifications aimed at maximising comparability of the information produced.

The minimum size of the sample of the overall population which is surveyed every year is of:

- Cross-sectional data operation: about 130,000 households and 270,000 persons aged 16 and more are interviewed in the European Union countries.
- Longitudinal data operation: about 100,000 households and 200,000 persons aged 16 and more are interviewed in the European Union countries.

The reference population in EU-SILC includes all private households and their current members residing in the territory of the countries at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population. Some small parts of the national territory amounting to no more than 2% of the national population and the national territories listed below may be excluded from EU-SILC. All household members are surveyed, but only those aged 16 and more are interviewed.
The construction of WIID3

The data points in a secondary database will originate from different sources and refer to a variety of income and population concepts, sample sizes, and statistical methods. To deal with this reality the only thing one can do is to specify as precisely as possible the conceptual base for each observation and to also otherwise document the data well. Atkinson & Brandolini (2001), Pyatt (2003), and Székeley & Hilgert (1999), who are critical of the use of secondary databases, point in particular to the problem of insufficient documentation. This criticism was taken into account in the construction of WIID2 (See the User Guide of WIID2, available at the UNU-WIDER web page).

In WIID3, we retain the basic strategy and structure of the earlier database, and try to report as thoroughly as possible the underlying data. The main changes with respect to WIID2 are the following:

New observations

The WIID3.3 version has now 7153 observations. The following summarizes the number of observations for different time periods:

Years:
Before 1970: 974
1970-89: 1,931
1990-99: 2,078
2000-13: 2,170

The new observations have been added from National Survey statistics obtained from the respective country official websites; the Socio-Economic Database for Latin America and the Caribbean (2012), Transmonee (2011), Luxembourg Income Study database, OECD, and EUROstat. Specific references are provided in the country documentation.

New observations have been added for a few Latin American countries from a new source, the Commitment to Equity Project, CEQ. This gives Gini’s both for Pre and Post Transfers for these countries (See Lustig and Higgins (2013) for the methodology and Lustig et al. 2013 for country-specific information).4

---

4 For more information, please also see http://www.commitmenttoequity.org/indicators.php
Corrected observations

The equivalence scale has been rationalized. Japan’s national data gives only the Elasticity Equivalent value and not the equivalence scale. However, \( \text{Equiv elasticity} = 0.5 \) is the square root scale and \( \text{Equiv elasticity} = 1 \) is the per capita scale. Hence to be consistent with the WIID methodology, the scale has been renamed for Japan.

In the case of SEDLAC data, the Equivalence scale used is not comparable to OECD modified scale or the square root method. Children under the age of 14 and between 14-18 are treated differently; hence it is called the SEDLAC scale.

Wherever Equivalence scale was missing, but the Unit of Analysis and Income Share unit was given, Equivalence scale has been derived and filled.

The variable IncDefn has been renamed as WelfareDefn. Welfare definition categories have been consolidated by correcting spellings etc.

The variables on Unit of analysis, Income share unit, and Equivalence scale have been reconciled. Family (Census or Economic or just Family) has been renamed as Household, and Income recipient has been renamed as Person for both Income Share unit and Unit of Analysis.

Most cases where mean and median incomes were given, but currency references were missing, have been corrected by referring to the source.

Some cases where the Gini values were unrealistically low/high have been corrected after checking from the source. Mean/Median value inconsistencies have been resolved to a large extent after cross-checking from the source.

**Gini** variable from WIID2c version has been dropped since the values obtained by using Shorrocks-Wan algorithm can now be computed using Stata command `ineqdeco`, after disaggregation using DASP utility.

A new variable called Revision has been added. This variable documents the changes made vis-à-vis the earlier databases.

Categorical variables indicating EU and OECD membership, regional classification, and time periods have been added.

Some duplicate observations and incorrect entries have been removed.

New variables

To assist the data visualization process, we have added variables containing classification of the income level of the countries from the World Bank, population and exchange rates. See the list of variables for details.

We now also offer simplified versions of some categorical variables, such as welfare definition and equivalence scales. The idea here is to combine categories.
which are close to each other conceptually to simplify choosing between different types of Gini. Again, these are especially useful for a query-based visualization system. See the Appendix on how the original variables (which are also retained) and the simplified variables are linked.

Format of the data base

The data are available in two formats, as an Excel file and as a Stata file. The dataset was prepared using Stata version 14, and the users of earlier version of the software need to do the following:

install -use13- by typing in Stata's command prompt:

ssc install use13

And then use the use13 command instead of the use command to open the data.

The documentation

The documentation of the database consists of three parts:

1) The documentation of the data in the database itself
2) This user guide
3) Country information sheets

The documentation in the database itself

In the database itself, the user is informed about the coverage of the surveys underlying the observations, the income sharing unit, the unit of analysis and the equivalence scale, the income concept and the source and survey used (for details on the variable please refer to the variable list below).

The following income/consumption/expenditure concepts are the ones that are mainly used:

Disposable income: This label is given if the income concept more or less corresponds to the one specified by the Canberra Group. Even if this label is given, some items might be badly covered. For example it is not always clear whether in-kind incomes are included or not. Often some in-kind incomes are covered but not home production. Sometimes non-labour incomes are asked in one question that lumps together transfers and income from property. The country-specific documentation and the quality rating give an indication if the income concept is acceptable.

Monetary disposable income: This label is given if there is a strong indication that in-kind incomes, imputed rents and home production are
not included and that the taxes are deducted from the incomes.

**Gross income:** This label is given if the income concept more or less corresponds to the one specified by the Canberra Group before the deduction of taxes and social contributions. The same comments as for the disposable incomes apply.

**Monetary gross income:** This label is given if there is a strong indication that in-kind incomes, imputed rents and home production are not included and that the taxes are not deducted from the incomes.

**Market income, factor income and primary income:** This label includes employee income, income from self-employment and property income. Market income also includes private pensions.

**Earnings** only refer to employee income and income from self-employment. A distinction between net and gross earnings has been made.

**Earnings** indicate that we do not know whether taxes have been deducted.

**Income:** This label is given if we do not have any information about the income concept from the source (or from some other sources). This means that the income concept might include earnings only, monetary incomes only, or it might be net or gross of taxes. Sources not including a definition of the income concept are accepted only if the source is one of the big income distribution compilations or if no other estimates are available for that country and year.

**Consumption:** This label is given if there is a strong indication that the use value, rather than the purchase value of durables is included or if durables are completely excluded. In addition, fines and taxes should not be included in the aggregation.

**Expenditure** This label is given if we know that durables are included with their purchase value and/or taxes and fines are included. This label is also given if we do not have information about the treatment of durables.

The following **income sharing units** are used:

**Household:** There are variations in the definitions. A broader definition defines the household as covering people who share a dwelling, a more restrictive definition those who share a dwelling and who share resources.

**Tax unit:** The definition depends on the tax laws but is often close to nuclear family. Sometimes children age 18 or over living with their parents are treated as separate tax units.

**Person:** Indicates that the data are collected on the individual level which is in general the case in earnings surveys.
The unit of analysis is either household or person. If the unit of analysis is household it means that the size of the households and the needs of different sized households have not been taken into account. If the unit is person it means that the needs of different sized households have been taken into account. The equivalence scale indicates that either no adjustment has been made for the difference in the relative need of different sized and composed households, or that an adjustment has been made. In the latter case the type of equivalence scale is indicated (for more general information about equivalence scales, please see the glossary). The country information sheets sometimes give more information about national equivalence scales. The four general scales that are used are:

- Household per capita: Household size
- Square root: Household size^{0.5}
- OECD scale: 1+0.7*n of additional adults + 0.5*n of children
- Modified OECD scale: 1+0.5*n of additional adults + 0.3*n of children

The country information sheets

In the country information sheets, we have summarized all the relevant documentation that has been available to us about the sources and the surveys used.

The sheets start by indicating the sources used and go on to describe the surveys. The years mentioned after the survey names indicate the years of the survey available to us, not the general availability of the survey. To understand the link between the country information sheets and the database it may be useful to check the variable Source Comments in the database. This column will in most cases indicate the name of the survey used for a particular estimate. The surveys indicated in this column are described in the sheets. We provide details about the survey coverage, sampling and income/consumption concepts, and if information was available on how the estimates were calculated in the source (column Source1 in the database), we also report that. The country information sheets will often give an impression of how consistent the time series are within sources and countries.

The quality rating

To give guidance in the use of the database, quality ratings were given to the observations. This was not an easy task because of the heterogeneity of the estimates and the difficulty to decide where to draw the line between high and low quality estimates. The lack of documentation for especially older observations is also a major problem.
The criteria used

We have used three criteria to evaluate the quality of a data point:

1) whether the concepts underlying the observations are known or not

In principle, this should be evident. In practice, it is far from always the case. Especially in older sources, it is often unclear what the income receiving units and the income concepts are.

2) the coverage of the income/consumption concept

The concepts as defined in the most preferred set of underlying definitions have been relied on (see table 1). For most developed countries, estimates based on monetary incomes have been accepted since the exclusion of in kind incomes and home production do not have a major effect on the income distribution. The exclusion of imputed rents does have some impact but since estimates are often not available, we have accepted the exclusion. In the case of earnings surveys, income concepts based on earnings are naturally accepted; in the case of household surveys not. This is because earnings do not give a complete picture of the household income. The exception is if the source reports estimates based on several different income concepts to illustrate the difference in inequality among different concepts. Deviations from the preferred income concept are if possible documented in the county information sheets.

3) the survey quality

A long list of desirable features could be pointed out, but in practice, coverage issues, questionnaires and data collection methodology were paid attention to. In many cases, the documentation available was insufficient to judge quality for even these issues. We often used additional sources to get information about the surveys.

Concerning coverage issues, we do not demand that the coverage should be national. Coverage is not necessarily a quality question, but about what is being measured. A rural household survey cannot be considered of bad quality because it covers rural areas only. The most important thing is that we know the survey coverage, so that rural or urban surveys are not taken for being national ones. Surveys covering very limited areas however are not acceptable, since they do not serve the purpose of the database. Attention was also paid to the exclusion of some special groups, such as households above a certain income threshold only living on charity.

Questionnaires or diaries need to have a sufficient level of income or expenditure detail to be acceptable.

The data collection methodology is especially important for expenditure surveys and in countries where a large proportion of the population works in the informal sector with infrequent incomes. In these cases, too long a recall period leads to considerable measurement errors. For expenditure
surveys, diaries must be kept or – especially in case of illiterate – frequent visits must be made to the households. Expenditure surveys collected in one single interview or with long recall periods were not considered to be of acceptable quality.

The final rating

These considerations resulted in the following quality rating:

1  (High quality) for observations
   a) where the underlying concepts are known
   b) where the quality of the income concept and the survey can be judged as sufficient according to the criteria described above

2  (Average quality) for observations where the quality of either the income concept or the survey is problematic or unknown or we have not been able to verify the estimates (the sources were not available to us); the country information sheets will often give an indication of the specific problems

3  (Low quality) for observations where both the income concept and the survey are problematic or unknown

4  for observations classified as memorandum items; some of the observations origin from the older compilations of inequality data have been given this rating since the data lying behind the observations often are unreliable

The interpretation of the quality rating should not be that only observations given rating 1 can be used. The other ones just do not satisfy the rather strict conditions that we have put up.
Some final guidelines

The user is advised to:

1) pay attention to definitional differences as documented in the database
2) consult the country sheets concerning information about individual countries
3) keep in mind that sources which adapt different income concepts or different statistical units cannot be combined or compared unless data corrections and adjustments are introduced
4) keep in mind that data points with similar definitions are not automatically comparable since differences in survey methodology might impair the comparability
5) report in their research paper which series of Ginis they used from the WIID; i.e. provide knowledge of their algorithms of data selection to make sure readers understand which observations were used.
List of Variables

- **Countrycode2** = 2-digit country code.
- **Countrycode3** = 3-digit country code.
- **Country** = country or area.
- **Year** (note that for a few observations for Estonia and Spain there are several quarterly observations for the same year, denoted in Survey/Source2 as Q1/Q2…)
- **Gini** coefficient as reported by the source. This replaces the ‘Reported Gini’ variable in WIID2.
- **Mean** = survey mean given with the same underlying definitions as the Gini coefficient and the share data.
- **Median** = survey median given with the same underlying definitions as the Gini coefficient and the share data.
- **Currency** = Gives the currency and the reference period for the means and medians. If the reference is US$90/month, it means that the currency is the 1990 US dollar per month. If the reference is US/month it means that the estimate is given in nominal value. Please note that in the visualized version of the data, all mean and median values are converted to annual ones and expressed in current USD values based on official exchange rates.
- **Reference_period** = time period for measuring mean and median incomes
- **Q1-Q5, D1-D10, P5, P95** = quintile, decile, percentile group shares.
- **AreaCovr** = area coverage. The land area which was included in the original sample surveys etc.
- **PopCovr** = population coverage. The population covered in the sample surveys in the land area (all, rural, urban etc.) which was included.
- **AgeCovr** = age coverage. Age limits imposed on the sample population. This is not explicitly given e.g. for the wage earning population, which – by definition – excludes children and most elderly people, unless special restrictions are used in the sample.
- **IncSharU** = income sharing unit/statistical unit.
- **UofAnala** = unit of analysis, indicates whether the data has been weighted with a person or a household weight.
- **Equivsc** = equivalence scale used.
- **Welfaredefn** = income/expenditure definition.
- **Source** = the source from which the observation value was obtained.
- **Source_Comments** = if the survey underlying the estimates is known this variable includes the name of the survey, otherwise it includes the source that Source1 cites as the (primary) source.
- **Revision** = Indicates the time of revision of the estimate. (1 = new observation in May 2007 revision, 2 = corrected in May 2007 revision, 3 = new observation in May 2008 revision, 4= corrected in May 2008 revision, 5= New Observation in 2014 revision)
- **Quality** = quality classification. (1 = high quality, 2 = average quality, 3 = low quality, 4 = not known)
- **Region** = regional groupings based on UN classification.
- **EU** = current EU member state
• OECD = current OECD member state
• Year_cat = decades (recent years), longer time spans (earlier years)
• Incomegroup: The World Bank classification by country income.⁵
• POP: Population of countries from Penn tables (World Bank)
• exchrate_penn: Exchange Rate (local) from Penn tables (World Bank)
• Welfaredefn_new: Simplified (recoded) welfare definition
• Equivse_new: Simplified (recoded) equivalence scale
• AgeCovr_new: Simplified (recoded) age coverage
• Area_Covr_new: Simplified (recoded) area coverage
• PopCovr_new: Simplified (recoded) population coverage
• IncSharU_new: Simplified (recoded) income sharing unit
• UofAnalysis_new: Simplified (recoded) unit of analysis

The Appendix provides more information on which categories have been combined in the generation of the last seven variables.

Glossary

The Lorenz Curve and the Gini Coefficient

---

A straightforward graphical interpretation of the Gini coefficient is in terms of the Lorenz curve, which is the thick curve in the figure above. The horizontal axis measures the cumulative percentage of the population, whose inequality is under consideration, starting from the poorest and ending with the richest. The vertical axis measures the cumulative percentage of income (or expenditure) associated with the units on the horizontal axis.

In case of a completely egalitarian income distribution in which the whole population has the same income, the Lorenz curve would be the dashed 45-degree line. When incomes vary within the population, the poor population has a proportionately lower share of income compared with the rich population, and the Lorenz curve may look like the above thick curve below the 45-degree line. As inequality rises, the thick curve moves towards the bottom right-hand corner.

The Gini coefficient is the area A between the 45-degree line and the Lorenz curve, divided by 1/2, the total area under the 45-degree line. The Gini coefficient may be given as a proportion or percentage. From this it is clear that the Gini coefficient will be equal to 0 when the distribution is equal. If the society's total income accrues to only one person/household unit, leaving the rest with no income at all, then the Gini coefficient approaches 1, or 100%.

Equivalence Scales
One complication posed by use of the household as the statistical unit is that households vary in size and composition and such differences between
households mean that their relative needs will be different. For example, a large household will have a lower standard of living from the same income as that received by a small household, all other things being equal. Costs of household members also differ according to their age, student status, labour force status and so on.

Equivalence scales are designed to adjust income/consumption to account for differences in need due to differences in household size and composition. The most basic of such adjustments is to calculate household income/consumption per member to adjust total incomes/consumption according to the number of people in the household. But such an adjustment ignores economies of scale in household consumption relating to size and other differences in needs among household members, in particular differing needs according to the age of both adults and children.

There is a wide range of equivalence scales in use in different countries and by different organisations. All take account of household or family size: in many scales this is the only factor, whilst in those taking into account other considerations it is the factor with greatest weight. Equivalence scales are usually presented as income/consumption amounts, or ratios of amounts, needed by households of different size and structure. Thus if a one person household needs one unit of income/consumption to maintain a given level of living, a two-person household may need 1.7 units, and a three-person household 2.2 units. There are two basic approaches to construction of scales: those which use the expert knowledge of social scientists and others, and those which are developed empirically based on analysis of survey data. (Citation from the Canberra Group Report, 2001, p.40)

**Quintile, decile, percentile group shares**

The quintile group shares express the share of total income going to each fifth of the population ordered according to the size of their incomes. In WIID2, these shares are expressed as percentages of total income. The first quintile group includes the poorest 20% of the population, while the fifth quintile includes the richest 20%. Deciles divide the population into ten groups and percentiles into one hundred groups.

**Unit record data / microdata**

Data that contain information on unit level from the survey; in the case of income or consumption distribution data the units is most often the household or the members of the household. If, for example, 8000 households took part in a survey, the unit record data include all 8000 households or household members.

**Grouped data**

This is data available in some kind of grouped form, for example the number of persons in income classes or quintile/decile group data.
Imputed rents for owner-occupied dwellings
This is the imputed value of the services provided by a household’s residence, after deduction of expenses, depreciation and property taxes. Home ownership may offset other costs and is therefore important. The main problem is the accurate measurement of imputed rent. The value of the rent of owner-occupied dwellings should in principle be the market rental value of an exactly similar house (Canberra Group Report, 2001, p.63 and p.120).

Home consumption
Value of goods produced and consumed within the households, less expenses incurred in production. Inclusion of this item is particularly important in countries where subsistence agriculture is significant (Canberra Group Report, 2001, p.120).
Appendix: Deriving simplified categorical variables

To introduce easy-to-use categorical variables, to complement the more detailed ones that we will still provide, we have done the following simplifications. The income and expenditure variable, named, “WelfareDefn” has more detailed categories (20) and “WelfareDefn_new “now has only the following 4 categories : i) Expenditure, ii) Gross income, iii) Disposable Income, and iv) others.

Column1 below has the detailed categories and Column 2 has consolidated categories.

<table>
<thead>
<tr>
<th>WelfareDefn</th>
<th>WelfareDefn_new</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption,</td>
<td>Consumption</td>
</tr>
<tr>
<td>Monetary Income (Gross), Market Income, Gross Income, Gross Earnings</td>
<td>Income, Gross</td>
</tr>
<tr>
<td>Income (Disposable), Monetary Income, (Disposable), Monetary Income, Disposable, (excl. self)</td>
<td>Income, Disposable</td>
</tr>
<tr>
<td>Monetary Income, Primary Income, Taxable Income, Factor Income, Income/Consumption</td>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EquivSc</th>
<th>EquivSc_new</th>
</tr>
</thead>
<tbody>
<tr>
<td>No adjustment</td>
<td>Without Adjustment</td>
</tr>
<tr>
<td>1988 Revised Jensen Scale , Census family eq, sqrt, Economic family eq, sqrt, Family eq OECD, Family eq . , Family eq, social asst, Family eq, square root, Family unit eq, national scale, Household eq, HBAI, Household eq, OECD</td>
<td>Household Adult Equiv</td>
</tr>
<tr>
<td>Family per capita, Household per capita, Per Capita</td>
<td>Household Per Capita</td>
</tr>
<tr>
<td>Tax unit eq, square root, Tax unit per capita, With equivalence elasticity = 0.5, With equivalence elasticity = 1</td>
<td>Other</td>
</tr>
<tr>
<td>AgeCovr</td>
<td>AgeCovr_new</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>10-, 15-, 16-, 17-, 20-, 20-64, 25-69</td>
<td>Age categories</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AreaCovr</th>
<th>AreaCovr_new</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apart from All, Rural and Urban categories, rest of the categories are clubbed under “Other”</td>
<td>Other</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Rural</td>
<td>Rural</td>
</tr>
<tr>
<td>Urban</td>
<td>Urban</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PopCov</th>
<th>PopCovr_new</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Rest all the other categories</td>
<td>Other</td>
</tr>
</tbody>
</table>

For Unit of Analysis and Income Share Unit, the categories of the recoded variable are the same as their original variable.
References


Belgium. *Statistical Yearbook.*


Denmark. *Danmarks Statistisk Arbog*. Various issues


European Commission (2013). Eurostat Database: Date of Extraction 23.5.2013

http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/main_tables


Luxembourg Income Study (LIS) (2013). Decile Estimates calculated by Markus Jantti of LIS provided to WIDER. The Gini’s have been extracted from :

http://www.lisdatacenter.org/data-access/key-figures/download-key-figures/


OECD: *Date of extraction 18.4.2014 from the following website*


Pakistan. *Statistical Yearbook*. Pakistan CSO.


45


Spain. Statistical Yearbook. Spain CSO.


Switzerland. *Statistical Yearbook 1999*. Switzerland CSO.


TransMONEE 2011 *A database of socio-economic indicators for CEE/CIS/Baltics*. Extracted on May 2013.,Table 10.8 and 10.9, from the following:

http://www.transmonee.org/


USA. US Bureau of the Census. Incomes [Internet]. Available from: <www.census.gov/>


49


