World Income Inequality Database

User Guide and Data Sources
## Contents

Preface................................................................................................................. 3
The basic principles behind WIID2................................................................. 4
  The conceptual base..................................................................................... 4
  Income or consumption?............................................................................. 4
  The income concept................................................................................... 4
  The consumption/expenditure concept...................................................... 5
  Other conceptual issues............................................................................. 7
The construction of WIID2.............................................................................. 7
  The database............................................................................................... 7
    Revision of the WIID1 data................................................................. 8
    New estimates added............................................................................. 8
    New variables added............................................................................. 9
    A new database building on earlier work........................................... 9
The documentation....................................................................................... 10
  The documentation in the database itself.............................................. 10
  The country information sheets.............................................................. 12
A revised quality rating.............................................................................. 12
  The quality rating in WIID1................................................................. 13
  The quality rating in D&S 1997............................................................... 13
The quality rating in WIID2....................................................................... 13
  The criteria used..................................................................................... 13
  The final rating....................................................................................... 14
Some final guidelines.................................................................................. 15
List of variables............................................................................................ 16
Glossary.......................................................................................................... 17
References...................................................................................................... 20
Preface

In the UNU/WIDER World Income Inequality Database (WIID) information on income inequality for developed, developing, and transition countries is stored.

WIID was initially compiled over 1997-1999 for the UNU/WIDER-UNDP project "Rising Income Inequality and Poverty Reduction: Are They Compatible?" directed by Giovanni Andrea Cornia, the former Director of UNU/WIDER. As more observations were added to the database, WIDER decided to make the database publicly available in order to facilitate further analysis and debate on inequality. This resulted in WIID version 1.0 which was published in September 2000. The database was designed by Renato Paniccia and Sampsa Kiiski, the programming was done by Sampsa Kiiski, and the data collected by Juha Honkkila, Renato Paniccia and Sampsa Kiiski.

The current update is part of the UNU/WIDER project "Global Trends in Inequality and Poverty" directed by Tony Shorrocks, the Director of UNU/WIDER and Guang Hua Wan, Senior Research Fellow at UNU/WIDER. The revision and the update were made by Susanna Sandström and the collection of old source material by Taina Iduozee. Markus Jäntti, professor at Åbo Akademi Univerity and Senior Research Associate at UNU/WIDER was the advisor for the revision.

We thank Klaus Deininger and Kihoon Lee from the World Bank and Lyn Squire from the Global Development Network for providing us with an update of the Deininger & Squire database, an update not published elsewhere. We are also grateful to Giovanni Andrea Cornia and his research assistant Luca Tiberti for providing us with their update of WIID. We thank the staff at UNICEF/ICDC in Florence for kindly providing us with additional details about the Transmonee data. Finally, several persons have contributed with data or comments: Peter Bolliger (Swiss Federal Statistical Office), Wim Bos (Statistics Netherlands), Andrea Brandolini (Bank of Italy), Kwang Soo Cheong (Johns Hopkins University), Jon Epland (Statistics Norway), Francisco J. Goerlich Gisbert (University of Valencia), Lee Rainwater (Harvard University and Luxembourg Income Study (retired)) and Timothy Smeeding (Maxwell Center for Policy Research and Luxembourg Income Study).
The basic principles behind WIID2

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 article on the use of secondary data in studies of income distribution, Atkinson & Brandolini (2001) 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 stone 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\(^1\), 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, preferably 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 can not 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\(^2\) 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.

---

\(^1\) Please refer to the glossary for an explanation of the terms used.
\(^2\) LSMS stands for Living Standards Measurement Study. The household surveys provided by this study can be found at [http://www.worldbank.org/lsms/](http://www.worldbank.org/lsms/).
Table 1 Preferred set of underlying concepts for inequality estimates in WIID2

<table>
<thead>
<tr>
<th>The income concept recommended by the Canberra Group for international comparisons of income distribution:</th>
<th>The consumption aggregate recommended by Deaton &amp; Zaidi (2002) for welfare measurements:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Employee income</strong>&lt;br&gt;Cash wages and salaries</td>
<td><strong>1. Food consumption</strong>&lt;br&gt;Food purchased from market&lt;br&gt;Home produced&lt;br&gt;Received as gift or in kind payment</td>
</tr>
<tr>
<td><strong>2. Income from self-employment</strong>&lt;br&gt;Profit/loss from unincorporated enterprise&lt;br&gt;Imputed income from self-employment&lt;br&gt;Goods and services produced for barter, less cost of inputs&lt;br&gt;Goods produce for home consumption, less cost of inputs</td>
<td><strong>2. Non-food consumption</strong>&lt;br&gt;Daily use items&lt;br&gt;Clothing and house wares&lt;br&gt;Health expenses&lt;br&gt;Education expenses&lt;br&gt;Transport</td>
</tr>
<tr>
<td><strong>3. Income less expenses from rentals, except rent of land</strong></td>
<td><strong>3. Durable goods</strong>&lt;br&gt;The use-value (rental value) of durables</td>
</tr>
<tr>
<td><strong>4. Property Income</strong>&lt;br&gt;Interest received less interest paid&lt;br&gt;Dividends</td>
<td><strong>4. Housing</strong>&lt;br&gt;Rents paid&lt;br&gt;If dwelling is owned by household or received free of charge, an estimate of the rental equivalent (imputed rent)&lt;br&gt;Utilities (water, electricity, garbage collection etc.)</td>
</tr>
<tr>
<td><strong>5. Current transfers received</strong>&lt;br&gt;Social insurance benefits from employers’ schemes&lt;br&gt;Social insurance benefits in cash from government schemes&lt;br&gt;Universal social assistance benefits in cash from government&lt;br&gt;Mean-tested social assistance benefits in cash from government&lt;br&gt;Regular inter-household cash transfers received</td>
<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/and taxes paid, purchase of assets, repayments of loans and lumpy expenditures, the concept to be referred to is expenditures.</td>
</tr>
<tr>
<td><strong>6. Total income (sum of 1 to 5)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>7. Current transfers paid</strong>&lt;br&gt;Employees’ social contributions&lt;br&gt;Taxes on income</td>
<td></td>
</tr>
<tr>
<td><strong>8. Disposable income (6 less 7)</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Other conceptual issues:**

1. **Household** should be the basic statistical unit
2. **Per capita** incomes or consumption/expenditure should be measured
3. **Person weights** should be applied
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 needs 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.

The construction of WIID2

The database

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

---

3 The Canberra Group mentions two common definitions of households: a broader definition covering people who share a dwelling and the more restrictive definition of those who share a dwelling and who usually eat together. Pyatt (2003) points out that the inclusion or exclusion of domestic servants, lodgers and absent family members can have a significant impact on the results in many developing countries.
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.

Revision of the WIID1 data

To address this source of criticism we have reviewed the data in the earlier database (as far as the sources have been available). Before any updating was done, WIID1 – including the data of Deninger & Squire 1997 (D&S 1997) – was first revised and cleaned. This was done to increase the preciseness of the definitions underlying the estimates and to update the documentation of the observations. WIID1 provided an excellent base to construct a well-documented database as its structure provided the infrastructure for a precise documentation of the conceptual base. To make WIID2 more user-friendly, overlapping estimates in WIID1, resulting from the merge of the estimates collected by WIDER and D&S 1997, were deleted. Low-quality estimates for country-years with high-quality estimates available were also deleted unless the estimates stemmed from one of the big compilations of inequality data. Almost all data points based on the Luxembourg Income Study (LIS) reported by D&S 1997, WIDER or other authors were deleted and replaced by new estimates, using our preferred definitions as outlined above, using the unit record data provided by LIS. The Transmonee data by UNICEF/ICDC were also re-entered, as updates to that source have been made. If several authors referred to the same source using very similar methods or referred to a source already included in the database, we strived to report only one estimate. This principle also led us to delete many estimates. Finally, estimates for very limited groups such as wage earners in metropolitan towns, were deleted if information on bigger population groups was available.

New estimates added

The new data of Deininger & Squire 2004 (D&S 2004), the unit record data of the Luxembourg Income Study (LIS), the Transmonee data by UNICEF/ICDC, Central Statistical Offices and research studies are central sources of the new estimates added in the database. The update by D&S 2004 is only published in WIID2 due to an agreement between the World Bank and WIDER to publish one database only. All the estimates of the new D&S are calculated by Kihoon Lee at the World Bank, using exclusively unit record data and using mostly our preferred definitions (the income and consumption concepts are sometimes different). From the LIS database, estimates not only based on the LIS disposable monetary income, but also on an extended concept including non-monetary incomes were calculated, if these income items were reported in the surveys.

---

4 Paukert (1973), Jain (1975), Cromwell (1977), UN (1981), Lecaillon et al. (1984), UN (1985) and Fields (1989) are considered as such.
5 For more information please look at http://www.lisproject.org/
6 For more information please look at http://www.unicef-icdc.org/resources/transmonee.html
7 In this case we also have been excluding estimates from bigger compilations like Jain 1975.
8 More precisely we added the LIS variables V3, V6, V9 and ALTNCASH to the LIS disposable income.
New variables added

We report two different Gini coefficients in WIID2. The first one is calculated by WIDER using methods developed by Tony Shorrocks and Guang Hua Wan to estimate the Gini coefficient from decile data almost as accurately as if unit record data were used (more information about the method will soon be made available on the website). When decile or quintile shares were not available, this Gini coefficient could not be estimated. The second Gini coefficient is called the “reported Gini” and is the one reported by the source or calculated by WIDER or Deininger & Squire for the old databases using POVCAL, a program for estimating the Gini coefficient using parametric extrapolation.

In WIID1, Gini coefficients and income shares of population groups were included in the data; in WIID2, survey estimates of means and medians are also included along with the income shares of the poorest 5 percent and richest 5 percent of the population whenever available. The survey means and medians are always based on the same definition as the Gini coefficients and the shares. If the Gini coefficients are based on per capita incomes, the means are reported per capita and if the Gini coefficients are based on household incomes, the means are reported for the households. The estimates are included as reported by the source and an additional variable indicates the currency and the reference period used. The decision to include survey means and medians was taken as these estimates at least in theory should provide and indication on the level of the living standards. They also give an indication about the quality of the survey since the means can be compared to the national accounts. In cases where both consumption- and income-based estimates are available, the mean can also give an indication of which of the two were measured more precisely. So far, few observations have data on the new variables. Our hope is that, as the database gets updated, more information on the new variables would be available. User response on the old database made us confident to publish the database in spreadsheet format. The Gini coefficients and income shares are not stored in two separate files as before but are all included in a single spreadsheet file.

A new database building on earlier work

The result of the revision process is that there is not a straightforward relationship between the old WIID1 and the new WIID2. Estimates have been deleted, exchanged with new updated ones and new estimates have been added. WIID1 was an excellent base to build on, but due to the criticism directed towards secondary databases (see Atkinson & Brandolini, 2001; Pyatt, 2003 and Székeley & Hilgert, 1999) we felt that a thorough revision was needed. WIID2 should therefore not be considered as an update of WIID1 but as a new database building on earlier work. One might argue that the update goes against the recommendation of Atkinson & Brandolini (2001) who emphasize that a secondary dataset should be a fully documented accumulation of earlier work, so that the user does not need to refer back to earlier datasets in order to obtain a

---

9 The importance of including mean incomes has also been stressed by Pyatt (2003).
complete picture of the available information. This is not the case, since only overlapping estimates and estimates that add no information have been deleted. The exception of this is that, as mentioned above, estimates included in the big compilations of income distribution have been kept even if they are of low quality (see footnote 4).

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 include 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.

Family is defined as a group of two or more persons residing together and related by birth, marriage, common-law or adoption. Whereas family refers to the nuclear family, economic family also allows other relatives to be present.

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 to column Survey/Source2 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.

A revised 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 quality rating for WIID2 was completely remade for all the estimates from the old WIID1, including the estimates in D&S 1997. This was
done since we felt that the quality ratings should reflect survey quality and income concepts. WIID1 and D&S 1997 also used a bit different criteria for their quality ratings, which mean that different estimates had been judged on different grounds when combined into a common database.

The quality rating in WIID1

In WIID1, the estimates were divided into “reliable” and “less reliable” based upon available information concerning area and population coverage, income recipient units, income concepts used, any survey design and sample size description included in the original sources, statistical year books, Deininger & Squire quality ratings, cross-references, and ad hoc information. Missing information, inconsistencies or large error possibilities in grouping or estimation methods, small population coverage, and generally limited data quality were the common reasons for excluding data points from the “reliable data” category. Due to limited availability of primary information, it was not possible to consistently control for survey composition, sampling methods, time period, non-response, weighting methods, minor differences in income definition and in-kind income adjustments, top coding and other statistical adjustments (WIDER, 2000).

The quality rating in D&S 1997

The D&S 1997 quality rating defined an “accept”-series using the three following criteria: 1) estimates had to be based upon survey data, 2) the survey had to have comprehensive population coverage (=national) and 3) a comprehensive measurement of income or expenditure (income inequality measures should include non-monetary incomes and not be based upon wage incomes only). A reexamination of the sources for D&S 1997 revealed several instances of mistakenly labeled “good quality estimates”, i.e., that did not, in fact, meet the criteria that had been set up.

The quality rating in WIID2

In the quality ratings of WIID2 the main principles as used in WIID1 and D&S 1997 are still adhered. The difference is that we have implemented quality ratings that differentiate the estimates into smaller categories, using to some extent the Canberra group criteria for such a classification.

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 can not 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 or households 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:
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

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

for observations where both the income concept and the survey are problematic or unknown

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

Compared to WIID1 and D&S 1997, we have included more categories to provide the user with more detailed information. By doing this we are able to separate the high quality estimates from ones giving a good indication of inequality but having some quality constraints either in the survey or the income concepts. Thus, 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
List of Variables

- **Country3** = 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 in percentage points as calculated by WIDER. If deciles or quintiles were not available this will be equal to the reported Gini.
- **Reported Gini** = the Gini as reported by the source (if no Gini were reported by the source, this will include the Gini as calculated by WIDER or Deininger & Squire for the old databases using POVCAL, a program estimating the Gini coefficient using parametric extrapolation).
- **Mean X/Y** = survey mean given with the same underlying definitions as the Gini coefficient and the share data.
- **Median X/Y** = survey median given with the same underlying definitions as the Gini coefficient and the share data.
- **Cur/ref** = 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.
- **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. This variable is corresponding to the variable sample unit in WIID1.
- **UofAnala** = unit of analysis, indicates wether the data has been weighted with a person or a household weight. This variable is corresponding to the variable enumeration unit in WIID1.
- **Equivsc** = equivalence scale used. This variable is corresponding to the variable reference unit in the old database.
- **IncDefn** = income/expenditure definition.
- **Source1** = the source from which the observation value was obtained.
- **Survey/Source2** = 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.
- **Quality** = quality classification.
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**

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**

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).
References


Belgium. Statistical Yearbook.


Chile (1994). *Notes on Income Distribution in Chile, 1994*. Ministerio de Planeacion (MIDEPLAN), Department of Planning and Social Studies.


28


Pakistan. *Statistical Yearbook*. Pakistan CSO.


Spain. Statistical Yearbook. Spain CSO.


Switzerland. *Statistical Yearbook 1999*. Switzerland CSO.


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


