World Income Inequality Database (WIID)

User guide and data sources

28 November 2023
Contents

Preface .................................................................................................................................................. iii
Basic principles behind the WIID ..................................................................................................... 4
  Conceptual base ............................................................................................................................... 4
  Income or consumption? .................................................................................................................... 4
  Income concept ................................................................................................................................. 4
  Consumption/expenditure concept ................................................................................................... 5
  Year ................................................................................................................................................ 5
Negative and zero values in the resource variables ........................................................................... 5
Other conceptual issues ....................................................................................................................... 6
Information regarding OECD, Eurostat, LIS, World Bank, ECLAC, and SEDLAC databases ........... 7
Evolution of the WIID ......................................................................................................................... 9
  Coverage ......................................................................................................................................... 10
  Corrected observations ..................................................................................................................... 10
  Approach to the grouped variables .................................................................................................. 10
  Variable level changes (since March 2021) .................................................................................... 11
  Database format ............................................................................................................................... 11
Documentation .................................................................................................................................. 11
  Documentation in the database itself ............................................................................................... 11
Quality rating ...................................................................................................................................... 14
  Criteria used ..................................................................................................................................... 14
  Final rating ....................................................................................................................................... 15
  Quality score ................................................................................................................................... 15
Some final guidelines ......................................................................................................................... 16
Referring to the WIID ......................................................................................................................... 16
List of variables ................................................................................................................................. 17
Glossary ............................................................................................................................................... 19
Appendix A: Codebook ...................................................................................................................... 21
Appendix B: United Nations geographical regions and sub-regions .................................................. 29
Appendix C: World Bank regional classification ............................................................................... 30
Sources ............................................................................................................................................... 31
References ......................................................................................................................................... 50
Preface

The UNU-WIDER World Income Inequality Database, better known as WIID, provides information on income inequality for 201 economies (including historical entities) in an organized and accessible manner.

The database has been an essential part of UNU-WIDER’s contribution to the study of inequalities within countries and globally since it was first launched in 2000, becoming one of the most prominent beacons of the institute’s identity. With inequality becoming one of the main megatrends of our time, and with the sustainable development goals explicitly aiming to reduce it, UNU-WIDER renews, with this new version, its commitment to making the best available data accessible.

The current version of WIID retains the main features of previous versions, especially the March 2021 version that extended its scope in many ways. First, by adding new measures of reported inequality, to better understand the complex distributional changes that can hardly be identified with a single index. Second, by producing companion datasets that facilitate the study of inequality within countries and, especially, at the global level. Third, by increasing the accessibility of this information to a wider range of potential users, especially non-experts. The data has been updated to incorporate new surveys from various countries. The number of observations has increased from 22,758 to 24,367.

As in previous versions, we strongly encourage comments and suggestions from all users to further improve the quality of WIID, both technically and in accessibility.

Kunal Sen

Director, UNU-WIDER

Helsinki, Finland
**Basic principles behind the WIID**

**Conceptual base**

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 and 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 and 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 and Bourguignon (2000) do not share this view. According to them, there is 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 and Zaidi 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 stepping stone in the construction of comparable statistics. In the WIID we have strived to collect observations with reference to both income and consumption, whenever it is possible.

**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, 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 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.

Consumption/expenditure concept

On the consumption side, the situation is more difficult. Deaton and Zaidi (2002) from the LSMS group at the World Bank 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 experiences, 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 lump expenditure 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.

Year

The year in the WIID typically refers to the year when the survey was finished. Note that in some cases this may be different from the recall period used for collecting information on income or consumption, which is used as the reference in some of the sources (e.g., PIP, LIS). Whenever there is not enough information, the WIID uses the year as in the original source.

Negative and zero values in the resource variables

Distributional values that are directly computed from microdata from LIS by the WIID team, are obtained after assigning a value arbitrarily close to zero (1 per cent of the mean) to income or

---

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: https://microdata.worldbank.org/index.php/catalog/lsms/about.
consumption values that are zero or negative in the microdata. This is done so all inequality measures can be estimated consistently with the minimum loss of information. In the rest of cases, the WIID reports the values as in the original source.

Other conceptual issues

The third issue to look at concerns relating to 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 wellbeing 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 because the existence of a lot of different equivalence scales weakens the comparability of the estimates.

c) *Person weights are preferred* as the users of income statistics are most often concerned with the economic well-being of individuals, 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 used in several cases. The differences appear especially in the statistical units and in the weighting.
Table 1: Preferred set of underlying concepts for inequality estimates in the WIID

<table>
<thead>
<tr>
<th>The income concept recommended by the Canberra Group for international comparisons of income distribution:</th>
</tr>
</thead>
<tbody>
<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>Means-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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The consumption aggregate recommended by Deaton and Zaidi (2002) for welfare measurements:</th>
</tr>
</thead>
<tbody>
<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 houseware</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/and taxes paid, purchase of assets, repayments of loans and lumpy expenditures, the concept to be referred to is expenditures.</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

Information regarding OECD, Eurostat, LIS, World Bank, ECLAC, and SEDLAC databases

WIID combines information coming from many sources, including historical compilations with updated information from the most salient data repositories (including LIS, ECLAC, SEDLAC, Eurostat, World Bank, and OECD), as well as from national statistical offices, and independent research papers. Below we introduce the main data sources.

**OECD**

The Organisation for Economic Co-operation and Development (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 standardized indicators based on the central concept of ‘equivalized household disposable income’; i.e., the total income received by 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 wellbeing, 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 maximize 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.

**Eurostat**

The European Union (EU) Statistics on Income and Living Conditions (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, Iceland, Norway, Switzerland, Montenegro, North Macedonia, Serbia, Turkey, and the United Kingdom:

- 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 harmonized 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 maximizing 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 over are interviewed in the European Union countries.
- Longitudinal data operation: about 100,000 households and 200,000 persons aged 16 and over 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 per cent 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 observations from Eurostat in the WIID are either estimated from microdata, whenever possible. The WIID also includes the series obtained from Eurostat website, with a larger coverage.

**LIS**

The Luxembourg Income Study (LIS) is the largest available income database of harmonized microdata and is based at the LIS Cross-National Data Center in Luxembourg. It mostly refers to
developed economies, but it is increasingly expanding to incorporate middle-income countries and, in the near future, also more low-income countries. This database is widely recognized as the main international reference for cross-country comparisons for the countries and years covered. The observations from LIS in the WIID are acquired through the institution’s LISSY remote-access system.

World Bank

The World Bank has provided an online tool, PovcalNet, to allow for country-level data estimation of inequality and poverty. The underlying concepts of the data acquired are somewhat difficult to track and thus we have graded the data mostly as average in our quality rating. Nevertheless this has been an important data source given its impressive coverage across countries.

In 2022, the World Bank announced that they were transitioning all their inequality related data to a new ‘Poverty and Inequality Platform – PIP’. We retain certain observations from the old PovcalNet, but most data points are now replaced by data from PIP.

ECLAC/UNECLAC

The United Nations Economic Commission for Latin America and the Caribbean (UNECLAC), often referred to as ECLAC or CEPAL (Comisión Económica para América Latina), hosts a wide range of statistics and indicators, accessible through the CEPALSTAT database. It is considered as one of the two main references for cross-country inequality comparisons in the region.

SEDLAC

The Socio-Economic Database for Latin America and the Caribbean (SEDLAC), based in CEDLAS (La Plata, Argentina) in collaboration with the World Bank, is a harmonized set of indicators based on a collection of surveys. It is considered as one of the two main references for cross-country inequality comparisons in the region. The WIID has acquired data directly from CEDLAS.

Evolution of the WIID

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 and Brandolini (2001), Pyatt (2003), and Székeley and 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 for WIID2, available from UNU-WIDER website). Jenkins (2015) provided a thorough review of WIID3 with suggestions on how it should be developed; Badgaiyan et al. (2015) addresses Jenkin’s comments in detail.

WIID4, released in 2018, included significant changes to the format of the database. Succeeding versions have mostly retained those concepts, but several new inequality measures have been since added to the database.

The WIID version launched in 2019 was the first iteration to no longer carry the version number.

In 2021, the WIID Companion was launched alongside the ‘traditional WIID’ for the first time.
Coverage

The WIID comprises of 22,758 observations. The following summarizes the number of observations for different time periods:

<table>
<thead>
<tr>
<th>Time span</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>24,367</td>
</tr>
<tr>
<td>Before 1960</td>
<td>311</td>
</tr>
<tr>
<td>1960–69</td>
<td>714</td>
</tr>
<tr>
<td>1970–79</td>
<td>946</td>
</tr>
<tr>
<td>1980–89</td>
<td>1,651</td>
</tr>
<tr>
<td>1990–99</td>
<td>3,758</td>
</tr>
<tr>
<td>2000–09</td>
<td>6,764</td>
</tr>
<tr>
<td>2010–19</td>
<td>8,753</td>
</tr>
<tr>
<td>2020–</td>
<td>1,470</td>
</tr>
</tbody>
</table>

The WIID contains data for 201 countries/economies, including historical entities, providing an almost worldwide coverage.

The observations originate from several sources:
1. LIS Cross-National Data Center (Luxembourg Income Study)
2. Eurostat
3. Socio-Economic Database for Latin America and the Caribbean (SEDLAC)
4. United Nations
5. Household survey statistics obtained from national statistical offices of the corresponding countries
6. Organisation for Economic Co-operation and Development (OECD)
7. World Bank’s Poverty and Inequality Platform (PIP)
8. Research outputs such as journal articles
9. Other international organizations

Corrected observations

WIID has been assembled from different sources, many dating back to times when paper records were the norm and transcription errors sometimes occurred. As a consequence, it included some duplicate observations, which have been eliminated, or coding errors and mistakes that have been corrected.

Some cases where the reported values of the Gini coefficient were inconsistent with historical trends have been verified with the source and corrected accordingly.

Approach to the grouped variables

The approach to the consolidated variables and the ‘full’ variables has been changed starting with WIID4. The main variable, also by name, is the grouped variable; e.g., scale, and then the additional information is given in the detailed variable; e.g., scale_detailed. It is convenient for most users to just use the consolidated variables.
For the grouped and detailed variables, it is convenient to follow which detailed values fall under which consolidated variable categories as the numerical values in the detailed variables are referring (with the first digit) to under which grouped category they fall into.

**Variable level changes (since March 2021)**

The Gini index variable is now named *gini*, it is still presented as in the originating source.

Several new inequality measures have been added to the database: the family of the generalized entropy measures, and the family of Atkinson’s inequality measures.

Variable *gdp* replaces the old variable *gdp_ppp_pc_usd2011*. The values represent gross domestic product (GDP) converted to 2017 US$ per capita (see Gradin 2021 for details).

**Database format**

The data are available in two formats, as an Excel file and as a Stata file. The dataset was prepared using Stata version 18.0, and the users of earlier version of the software may need to do the following:

```stata
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.
```

**Documentation**

The documentation of the database consists of four parts:

1. the documentation of the data in the database itself
2. this user guide
3. technical note paper
4. country information sheets

**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:

*Net income/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**: This label only refer to employee income and income from self-employment. A distinction between net and gross earnings has been made. *Earnings* (without a notion of gross or net) indicates 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.

It is important to note that the distinction between gross and net incomes is sometimes problematic. For example, this is a well-known problem in many Latin American surveys. The issue is that some questionnaires tend to implicitly request gross income, while there is the belief that people paying direct taxes (the formal sector) might actually be reporting take-home wages. For the informal sector, there is basically no difference between net and gross. For this reason, in some cases in which this problem is identified and the source is not clear about whether income is gross or net, income is labelled as net/gross to indicate this ambiguity. In other cases, it is possible that income, even if labelled as net or as gross, still has the same problem.
The following *income sharing units* (variable *sharing_unit*) are used (mainly):

*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 aged 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* (variable *reference_unit*) 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* (variable *scale*) captures the way in which the resource levels of economic units are converted into the resource levels of the population units when equal sharing is assumed. ‘No adjustment’ is recorded when the population unit is the same as the economic unit. But various options are possible when, as is often the case, the original data refer to households, but the desired income distribution is defined over individuals.

If household needs rise in proportion to household size, then it is appropriate to assign household resources per capita to each household member, assuming equal sharing, as is frequently done in the WIID data. At the other extreme, making no adjustment and assigning total household resources to each household member implicitly assumes that additional household members do not increase needs (and there is equal sharing again). Empirical evidence suggests an intermediate position — household needs rise with size, but not in proportion due to economies of scale in consumption. Household equivalence scales reflect this evidence, but differ across time and place, perhaps reflecting differences in household technology and spending patterns but also no doubt due to estimation methods. Thus, there are many different equivalence scales.

The four main general scales that are used are:

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

If the variation in equivalence scales used in different cases reflects genuine differences in technology or spending patterns, then that variation is not a concern. But even if the variation does not have such a justification, from the viewpoint of WIID users, the multiplicity of equivalence scales is probably a distraction rather than useful information. So, we group them all together and distinguish only three categories for the scale variable: per capita, equivalized, or no adjustment. Note that if a per capita or equivalized scale is applied to household resources, then the population unit must be the individual.

13
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.

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 illiteracy, 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.

Final rating

These considerations resulted in the following quality rating:

- **High quality** refers to observations where both (a) the underlying income or consumption concepts are known and (b) the quality of the income or consumption concept and the survey are satisfactory according to the criteria outlined above.
- **Average quality** refers to observations where either (a) the underlying income or consumption concept or else (b) the quality of the income concept and the survey are unknown or unsatisfactory. The country information sheets will often indicate the specific problems.
- **Low quality** indicates observations where both the income or consumption concept and the survey quality are unsatisfactory.
- **Not known** is the label we attach to observations for which income concept and the survey quality are both indeterminate due to insufficient information. This rating is more common for older observations due to poor documentation.

Note that the quality assessment is intended as guidance for users, not as a recommendation that users discard observations not judged to be high quality. While the other observations do not satisfy the rather strict conditions that we have applied, they will still be useful in most applications.

Quality score

In addition to the quality variable, we provide also a computed quality score. This aims at giving a sense of how much information is provided by each observation, under the understanding that the more information we have about the survey and methodology used to produce the estimates, the better. It also considers how close the estimates are from the standard ones used in the literature. It does not make any consideration, however, about the quality of the survey or the methodology.

We award points to the observations based on their attributes in the following way (maximum is 13 points)

- Gini coefficient is available (1)
- Resource concept:
  - Consumption, Income (net), Income (gross), Monetary income (gross), Monetary income (net) (5)
  - Income, Income (net/gross), Monetary income, Market income (3)
  - Factor income, Primary income, Taxable income, Earnings (1)
- Equivalence scale:
  - Per capita or equivalized (3)
  - No adjustment (2)
- Area coverage:
  - All, Urban, Rural (1)
- Population coverage:
  - All (1)
- Distributional share information:
  - All of \( d1-d1 \) are available (2)
  - All of \( q1-q5 \) are available (at least one of \( d1-d10 \) is missing) (1)
Some final guidelines

When using this WIID main dataset, the user is advised to:
1. pay attention to definitional differences as documented in the database
2. 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
3. keep in mind that data points with similar definitions are not automatically comparable since differences in survey methodology might impair the comparability
4. 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

For a more comparable dataset in which the concepts have been standardised and are consistent across countries and over time, the user is advised to use the WIID Companion datasets instead.

Referring to the WIID

Please refer to the present WIID as:
https://doi.org/10.35188/UNU-WIDER/WIID-281123
# List of variables

## Variables used in the WIID

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Identifier</td>
</tr>
<tr>
<td>country</td>
<td>Country or area</td>
</tr>
<tr>
<td>c3</td>
<td>3-digit country code in ISO 3166-1 alpha-3 format</td>
</tr>
<tr>
<td>c2</td>
<td>2-digit country code in ISO 3166-1 alpha-2 format</td>
</tr>
<tr>
<td>year</td>
<td>Year. This typically refers to the survey year.</td>
</tr>
<tr>
<td>gini</td>
<td>Gini coefficient</td>
</tr>
<tr>
<td>ge0</td>
<td>GE(0), MLD, M-Theil</td>
</tr>
<tr>
<td>ge1</td>
<td>GE(1), T-Theil</td>
</tr>
<tr>
<td>ge2</td>
<td>GE(2), $\frac{1}{2} CV^2$</td>
</tr>
<tr>
<td>a025</td>
<td>Atkinson (0.25)</td>
</tr>
<tr>
<td>a050</td>
<td>Atkinson (0.50)</td>
</tr>
<tr>
<td>a075</td>
<td>Atkinson (0.75)</td>
</tr>
<tr>
<td>a1</td>
<td>Atkinson (1)</td>
</tr>
<tr>
<td>palma</td>
<td>Palma ratio -- the share of the top 10% divided by the share of the bottom 40%</td>
</tr>
<tr>
<td>ratio_top20bottom20</td>
<td>The share of the top 20% divided by the share of the bottom 20%</td>
</tr>
<tr>
<td>bottom40</td>
<td>Bottom 40%, share of the total. Evolution of this measure over time is one of the indicators of the Sustainable Development Goals.</td>
</tr>
<tr>
<td>q1-q5</td>
<td>Quintile group shares of resource</td>
</tr>
<tr>
<td>d1-d10</td>
<td>Decile group shares of resource</td>
</tr>
<tr>
<td>bottom5 and top5</td>
<td>Bottom five and top 5% group shares of resource</td>
</tr>
<tr>
<td>resource</td>
<td>Resource concept</td>
</tr>
<tr>
<td>resource_detailed</td>
<td>Detailed resource concept</td>
</tr>
<tr>
<td>scale</td>
<td>Equivalence scale</td>
</tr>
<tr>
<td>scale_detailed</td>
<td>Detailed equivalence scale</td>
</tr>
<tr>
<td>sharing_unit</td>
<td>Income sharing unit/statistical unit</td>
</tr>
<tr>
<td>reference_unit</td>
<td>Unit of analysis, indicates whether the data has been weighted with a person or a household weight</td>
</tr>
<tr>
<td>areacovr</td>
<td>Area coverage. The land area which was included in the original sample surveys etc.</td>
</tr>
<tr>
<td>areacovr_detailed</td>
<td>Detailed area coverage</td>
</tr>
<tr>
<td>popcovr</td>
<td>Population coverage. The population covered in the sample surveys in the land area (all, rural, urban etc.) which was included</td>
</tr>
<tr>
<td>popcovr_detailed</td>
<td>Detailed population coverage, including age coverage information in certain cases</td>
</tr>
<tr>
<td>region_un</td>
<td>Regional grouping based on United Nations geoscheme</td>
</tr>
</tbody>
</table>
**region_un_sub** | Sub-regional grouping based on United Nations geoscheme

**region_wb** | Regional grouping based on World Bank classification

**eu** | Current EU member state

**oecd** | Current OECD member state

**incomegroup** | World Bank classification by country income

**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** | Currency for the mean and median values. If the reference is US$2011PPP it means that the currency is in 2011 US dollar per month, with purchasing power parity applied on it.

**reference_period** | Time period for measuring mean and median values

**exchangerate** | Conversion rate from local currency units (LCU) to United States Dollars (USD)

**mean_usd** | Mean measure in United States Dollar (USD)

**median_usd** | Median measure in United States Dollar (USD)

**gdp** | Gross domestic product (GDP) converted to 2017 US$ in per capita terms, integrated series

**population** | Population of countries from the UN population prospects

**revision** | Indicates the time of the revision when the observation was included to the database

**quality** | Quality assessment

**quality_score** | Computed quality score

**source** | Source type

**source_detailed** | Source from which the observation was obtained

**source_comments** | Additional source comments

**survey** | Originating survey information

**link** | Link to the source at the time of extracting the data

**wiidcompanion** | Indicates whether the observation was used to construct the country companion dataset
Glossary

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 organizations. 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: 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 the WIID, these shares are expressed as percentages of total income. The first quintile group includes the poorest 20 per cent of the population, while the fifth quintile includes the richest 20 per cent. 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, 8,000 households took part in a survey, the unit record data include all 8,000 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: 63, 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: 120).
Appendix A: Codebook

The value labels for numeric variables are listed in below. In addition to the grouped variables (e.g., \textit{resource}), we provide the detailed variables (e.g., \textit{resource\_detailed}) that contain the full information. The values of the detailed variables match to the grouped variables with the first digit: For example, ‘202 Monetary income’ in the \textit{resource\_detailed} is ‘2 Income (net/gross)’ in \textit{resource}.

<table>
<thead>
<tr>
<th>\textit{resource}</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Income (net)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Income (net/gross)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Income (gross)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Consumption</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Earnings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\textit{resource_detailed}</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>101</td>
<td>Income, net</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>Monetary income, net</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>Monetary income, net (excluding property income)</td>
</tr>
<tr>
<td></td>
<td>201</td>
<td>Income, net/gross</td>
</tr>
<tr>
<td></td>
<td>202</td>
<td>Monetary income</td>
</tr>
<tr>
<td></td>
<td>301</td>
<td>Income, gross</td>
</tr>
<tr>
<td></td>
<td>302</td>
<td>Monetary income, gross</td>
</tr>
<tr>
<td></td>
<td>401</td>
<td>Consumption</td>
</tr>
<tr>
<td></td>
<td>501</td>
<td>Earnings</td>
</tr>
<tr>
<td></td>
<td>502</td>
<td>Earnings, gross</td>
</tr>
<tr>
<td></td>
<td>503</td>
<td>Earnings, net</td>
</tr>
<tr>
<td></td>
<td>504</td>
<td>Factor income</td>
</tr>
<tr>
<td></td>
<td>505</td>
<td>Market income</td>
</tr>
<tr>
<td></td>
<td>506</td>
<td>Primary income</td>
</tr>
<tr>
<td></td>
<td>507</td>
<td>Taxable income, excluding property income</td>
</tr>
<tr>
<td></td>
<td>508</td>
<td>Taxable income, gross</td>
</tr>
<tr>
<td></td>
<td>509</td>
<td>Taxable income, gross (including deductions)</td>
</tr>
<tr>
<td></td>
<td>510</td>
<td>Taxable income, net</td>
</tr>
<tr>
<td></td>
<td>601</td>
<td>Income/consumption</td>
</tr>
<tr>
<td><strong>scale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>Label</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Per capita</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Equivalized</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>No adjustment</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>scale_detailed</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Label</td>
</tr>
<tr>
<td>101</td>
<td>Per capita</td>
</tr>
<tr>
<td>102</td>
<td>Head of household</td>
</tr>
<tr>
<td>201</td>
<td>Equivalized</td>
</tr>
<tr>
<td>202</td>
<td>1977 McClements scale</td>
</tr>
<tr>
<td>203</td>
<td>1988 revised Jensen scale</td>
</tr>
<tr>
<td>204</td>
<td>National scale</td>
</tr>
<tr>
<td>205</td>
<td>OECD</td>
</tr>
<tr>
<td>206</td>
<td>OECD-modified</td>
</tr>
<tr>
<td>207</td>
<td>Square root</td>
</tr>
<tr>
<td>208</td>
<td>Supplemental poverty measure</td>
</tr>
<tr>
<td>209</td>
<td>SEDLAC</td>
</tr>
<tr>
<td>301</td>
<td>No adjustment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>sharing_unit</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Label</td>
</tr>
<tr>
<td>1</td>
<td>Household</td>
</tr>
<tr>
<td>2</td>
<td>Family</td>
</tr>
<tr>
<td>3</td>
<td>Tax unit</td>
</tr>
<tr>
<td>4</td>
<td>Person</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>reference_unit</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Label</td>
</tr>
<tr>
<td>1</td>
<td>Person</td>
</tr>
<tr>
<td>2</td>
<td>Household</td>
</tr>
<tr>
<td>3</td>
<td>Family</td>
</tr>
<tr>
<td>4</td>
<td>Tax unit</td>
</tr>
</tbody>
</table>
### areacovr

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
</tr>
<tr>
<td>2</td>
<td>Rural</td>
</tr>
<tr>
<td>3</td>
<td>Urban</td>
</tr>
<tr>
<td>4</td>
<td>Part</td>
</tr>
</tbody>
</table>

### areacovr_detailed

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>All</td>
</tr>
<tr>
<td>102</td>
<td>All, excl. Abkhasia and Tsikhinvali</td>
</tr>
<tr>
<td>103</td>
<td>All, excl. Costa Rural, Selva Rural and Selva Urbana (30% of the population)</td>
</tr>
<tr>
<td>104</td>
<td>All, excl. East Timor</td>
</tr>
<tr>
<td>105</td>
<td>All, excl. East-Central State</td>
</tr>
<tr>
<td>106</td>
<td>All, excl. Transnistria</td>
</tr>
<tr>
<td>107</td>
<td>All, excl. West Irian and East Timor</td>
</tr>
<tr>
<td>108</td>
<td>All, excl. West Irian, East Timor and Maluku</td>
</tr>
<tr>
<td>109</td>
<td>All, excl. eight districts in the north and the east (15% of the population)</td>
</tr>
<tr>
<td>110</td>
<td>All, excl. nomadic areas</td>
</tr>
<tr>
<td>111</td>
<td>All, excl. northern and eastern provinces</td>
</tr>
<tr>
<td>112</td>
<td>All, excl. seven districts</td>
</tr>
<tr>
<td>113</td>
<td>All, excl. some special areas (4% of the population)</td>
</tr>
<tr>
<td>114</td>
<td>Continental Portugal</td>
</tr>
<tr>
<td>115</td>
<td>Main island</td>
</tr>
<tr>
<td>116</td>
<td>With rural north</td>
</tr>
<tr>
<td>117</td>
<td>Without rural north</td>
</tr>
<tr>
<td>118</td>
<td>Without Northern Ireland</td>
</tr>
<tr>
<td>201</td>
<td>Rural</td>
</tr>
<tr>
<td>202</td>
<td>Agricultural sector</td>
</tr>
<tr>
<td>203</td>
<td>Four rural areas</td>
</tr>
<tr>
<td>204</td>
<td>Rural, excl. seven districts on national level</td>
</tr>
<tr>
<td>301</td>
<td>Urban</td>
</tr>
<tr>
<td>302</td>
<td>All, mainly urban areas</td>
</tr>
<tr>
<td>303</td>
<td>Capital</td>
</tr>
<tr>
<td>304</td>
<td>Cities</td>
</tr>
<tr>
<td>305</td>
<td>Cities (n=16)</td>
</tr>
<tr>
<td>306</td>
<td>Cities (n=17)</td>
</tr>
<tr>
<td>307</td>
<td>Cities (n=4)</td>
</tr>
<tr>
<td>308</td>
<td>Cities (n=7)</td>
</tr>
<tr>
<td>309</td>
<td>Cities (n=8)</td>
</tr>
<tr>
<td>310</td>
<td>Metropolitan area</td>
</tr>
<tr>
<td>311</td>
<td>Nonagricultural sector</td>
</tr>
<tr>
<td>312</td>
<td>Paramaribo and Wanica</td>
</tr>
<tr>
<td>313</td>
<td>Urban, excl. Western Province</td>
</tr>
<tr>
<td>Value</td>
<td>Label</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>All</td>
</tr>
<tr>
<td>2</td>
<td>Economically active</td>
</tr>
<tr>
<td>3</td>
<td>Specific categories</td>
</tr>
</tbody>
</table>

### popcovr

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>All</td>
</tr>
<tr>
<td>201</td>
<td>Economically active</td>
</tr>
<tr>
<td>202</td>
<td>Employed</td>
</tr>
<tr>
<td>203</td>
<td>Family units with earnings</td>
</tr>
<tr>
<td>204</td>
<td>Households with earnings</td>
</tr>
<tr>
<td>205</td>
<td>Income recipients</td>
</tr>
<tr>
<td>206</td>
<td>Taxpayers</td>
</tr>
<tr>
<td>301</td>
<td>Agricultural households</td>
</tr>
<tr>
<td>302</td>
<td>All excl. some private sector employees</td>
</tr>
<tr>
<td>303</td>
<td>All, aged 20-64</td>
</tr>
<tr>
<td>304</td>
<td>All, aged 25-59</td>
</tr>
<tr>
<td>305</td>
<td>All, excl. farmers</td>
</tr>
<tr>
<td>306</td>
<td>All, excl. fishery hhs and farm hhs with very small land holdings</td>
</tr>
<tr>
<td>307</td>
<td>All, excl. foreign-head hhs + hhs with net income DM &gt; 25000</td>
</tr>
<tr>
<td>308</td>
<td>All, excl. foreign-headed hhs</td>
</tr>
<tr>
<td>309</td>
<td>All, excl. foreign-headed hhs + hhs with net income DM &gt;= 15000</td>
</tr>
<tr>
<td>310</td>
<td>All, excl. foreign-headed hhs + hhs with net income DM &gt;= 20000</td>
</tr>
<tr>
<td>311</td>
<td>All, excl. foreign-headed hhs + hhs with net income over a certain limit</td>
</tr>
<tr>
<td>312</td>
<td>All, excl. hhs with net income DM &gt; 35000</td>
</tr>
<tr>
<td>313</td>
<td>All, excl. hhs with wives aged 44+</td>
</tr>
<tr>
<td>314</td>
<td>All, excl. households depending entirely on charity</td>
</tr>
</tbody>
</table>
315 | All, excl. nomadic people (30% of the population)  
316 | All, excl. pensioner-headed households  
317 | All, excl. pensioners  
318 | All, excl. self-employeds in the high income brackets  
319 | All, excl. single-member households  
320 | All, excl. very high income households  
321 | All, unclear if inclusive of nomadic people (30% of the population)  
322 | Employed, > 10 employees  
323 | Employed, > 25 employees  
324 | Employed, > 5 employees  
325 | Employed, African males  
326 | Employed, aged 10+  
327 | Employed, aged 16+  
328 | Employed, excl. entrepreneurs and farmers, >= 3 employees  
329 | Employed, excl. independent farmers, persons employed in crafts and trade  
330 | Employed, excl. private enterprises, >= 20 employees  
331 | Employed, excl. self-employeds  
332 | Employed, excl. small enterprises  
333 | Employed, excl. small enterprises and cooperatives  
334 | Employed, excl. small private enterprises  
335 | Employed, full-time  
336 | Employed, full-time employees in the public sector  
337 | Employed, full-time, >= 100 employees  
338 | Employed, full-time, >= 20 employees  
339 | Employed, full-time, >= 25 employees  
340 | Employed, full-time, >= 25 employees, some sectors >= 100 employees  
341 | Employed, full-time, >= 50 employees  
342 | Employed, full-time, excl. self-employeds and farmers  
343 | Employed, multi-member households  
344 | Employed, private sector  
345 | Employed, public sector  
346 | Employed, public sector, excl. social organizations  
347 | Employed, socialized sector  
348 | Employed, socialized sector, > 5 employees  
349 | Employed, state and cooperative sector  
350 | Employed, state sector  
351 | Employee households  
352 | Estate sector  
353 | Excl. self-employed households  
354 | Households where head employed or inactive  
355 | Households with positive or zero taxable incomes  
356 | Income recipients, aged 17+  
357 | Income recipients, public sector  
358 | Males, aged 20+  
359 | Mostly families of state sector and collective farm employees
<table>
<thead>
<tr>
<th></th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>360</td>
<td>Non-agricultural households</td>
</tr>
<tr>
<td>361</td>
<td>Non-agricultural multi-member households</td>
</tr>
<tr>
<td>362</td>
<td>Non-estate sector</td>
</tr>
<tr>
<td>363</td>
<td>Non-farm population</td>
</tr>
<tr>
<td>364</td>
<td>Omani</td>
</tr>
<tr>
<td>365</td>
<td>Qatari</td>
</tr>
<tr>
<td>366</td>
<td>Saudi</td>
</tr>
<tr>
<td>367</td>
<td>Self-employed households</td>
</tr>
<tr>
<td>368</td>
<td>Taxpayers, Jewish</td>
</tr>
<tr>
<td>369</td>
<td>Taxpayers, aged 15+</td>
</tr>
<tr>
<td>370</td>
<td>Taxpayers, aged 16+</td>
</tr>
<tr>
<td>371</td>
<td>Taxpayers, aged 20+</td>
</tr>
<tr>
<td>372</td>
<td>Taxpayers, permanently employed and self-employed</td>
</tr>
<tr>
<td>373</td>
<td>Wage earners</td>
</tr>
<tr>
<td>374</td>
<td>Workers</td>
</tr>
<tr>
<td>375</td>
<td>Workers, state and cooperative sector</td>
</tr>
</tbody>
</table>

### region_un

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Americas</td>
</tr>
<tr>
<td>2</td>
<td>Europe</td>
</tr>
<tr>
<td>3</td>
<td>Africa</td>
</tr>
<tr>
<td>4</td>
<td>Asia</td>
</tr>
<tr>
<td>5</td>
<td>Oceania</td>
</tr>
</tbody>
</table>

### region_un_sub

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Northern America</td>
</tr>
<tr>
<td>102</td>
<td>Central America</td>
</tr>
<tr>
<td>103</td>
<td>Caribbean</td>
</tr>
<tr>
<td>104</td>
<td>South America</td>
</tr>
<tr>
<td>201</td>
<td>Northern Europe</td>
</tr>
<tr>
<td>202</td>
<td>Western Europe</td>
</tr>
<tr>
<td>203</td>
<td>Eastern Europe</td>
</tr>
<tr>
<td>204</td>
<td>Southern Europe</td>
</tr>
<tr>
<td>301</td>
<td>Northern Africa</td>
</tr>
<tr>
<td>302</td>
<td>Western Africa</td>
</tr>
<tr>
<td>303</td>
<td>Middle Africa</td>
</tr>
<tr>
<td>304</td>
<td>Eastern Africa</td>
</tr>
<tr>
<td>305</td>
<td>Southern Africa</td>
</tr>
<tr>
<td>401</td>
<td>Western Asia</td>
</tr>
<tr>
<td>402</td>
<td>Central Asia</td>
</tr>
<tr>
<td>403</td>
<td>Southern Asia</td>
</tr>
<tr>
<td>404</td>
<td>Eastern Asia</td>
</tr>
<tr>
<td>region_wb</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>eu</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Non-EU</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>EU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>oecd</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Non-OECD</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>OECD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>incomegroup</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>High income</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Upper middle income</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Lower middle income</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Low income</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>reference_period</th>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Year</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Month</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Week</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Day</td>
</tr>
</tbody>
</table>
### quality

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>Average</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>Not known</td>
</tr>
</tbody>
</table>

### source

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Luxembourg Income Study</td>
</tr>
<tr>
<td>2</td>
<td>Eurostat</td>
</tr>
<tr>
<td>3</td>
<td>SEDLAC</td>
</tr>
<tr>
<td>4</td>
<td>United Nations</td>
</tr>
<tr>
<td>5</td>
<td>National statistical authority</td>
</tr>
<tr>
<td>6</td>
<td>OECD</td>
</tr>
<tr>
<td>7</td>
<td>World Bank</td>
</tr>
<tr>
<td>8</td>
<td>Research study</td>
</tr>
<tr>
<td>9</td>
<td>Other international organizations</td>
</tr>
</tbody>
</table>

### wiidcompanion

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No (overlapping)</td>
</tr>
</tbody>
</table>
Appendix B: United Nations geographical regions and sub-regions
Appendix C: World Bank regional classification
Sources


Hong Kong (various). Population Census. Hong Kong: Census and Statistics Department.


LIS Cross-National Data Center (Luxembourg Income Study) (various). Distributional income data received from LIS.


SEDLAC (CEDLAS and The World Bank) (various). *Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank)*. CEDLAS Universidad Nacional de La Plata and The World Bank’s LAC Poverty Group (LCSPP) and MECOVI Program.


Switzerland. *Statistical Yearbook 1999*. Switzerland CSO.


Taiwan (various). *Survey of Personal Income Distribution*. Taipei: Taiwan Directorate General of Budget, Accounting and Statistics.

Taiwan (various). *Survey of Family Income & Expenditure*. Taipei: Taiwan Directorate General of Budget, Accounting and Statistics.


Referencias


