
Note on Income Inequality Data

Global Dialogue for Social Development Branch

Emerging Issues and Trends Section

October 2018



Note on income inequality data

There has been much progress in improving the availability, quality and comparability of income and wealth inequality data. Several cross-national databases containing summary inequality statistics are now available. In this note, we review the World Bank's PovcalNet, the Luxembourg Income Study and Wealth Study Databases (LIS, LWS), the Standardized World Income Inequality Database (SWIID), the World Income Inequality Database (WIID), the World and Wealth Income Database (WID.world), All the Ginis dataset, the Estimated Household Income Inequality dataset (EHII) and the Global Consumption and Income Project (GCIP).¹ All data are publicly available free of charge in all the databases examined with the exception LIS/LWS.

The databases reviewed differ considerably in purpose, coverage, data sources and indicators provided. Some of them are just repositories of estimates compiled from primary and other secondary sources. Others provide original estimates based on microdata from, mainly, a growing number of household surveys. Some rely on imputation methods to obtain estimates for years when data are missing while others do not. As a result, coverage by country and year differs significantly across datasets. Some databases are produced by institutions while others are developed by individual researchers. Some institutions make data harmonization one of their priorities while others offer diverse sets of data—and the metadata needed to identify differences across data sources and countries.

Although there is significant agreement among these datasets, there are also inconsistencies in both the levels and trends of inequality obtained from each database (for each given indicator). Some of the differences across databases are illustrated below.

Overall, there are trade-offs between breadth (coverage) and comparability. Maximizing comparability and quality means focusing on a small number of (developed) countries. It also requires thoroughly harmonizing data, using data from one source or using only a single basis of calculation. Increasing coverage means relying on less reliable data, using different variables to produce estimates (income is used in practically all developed countries; consumption is often the underlying measure in developing countries), and/or making assumptions to impute values where data are missing.

Among the databases examined, PovcalNet appears to have the most non-imputed estimates for the largest number of countries. It is also the data source used for the international monitoring of SDG target 10.1.² On the other hand, LIS is the only source that uses a uniform set of assumptions and definitions on the basis of thoroughly harmonized microdata to maximize comparability. SWIID provides the most complete dataset, but many of the values are imputed.

¹ The note does not include regional databases such as those available for Latin American countries (e.g. those generated by CEDLAS and ECLAC) as well as the OECD Income Distribution Database. The note does not include, either, datasets exclusively on earnings or other specific components of income.

² By 2030, progressively achieve and sustain income growth of the bottom 40 per cent of the population at a rate higher than the national average.



Databases for the measurement of income inequality

I. PovcalNet/World Development Indicators (WDI)

Host institution: World Bank.

Coverage (number of countries and years): More than 1,000 household surveys from 159 countries, although coverage varies depending on the year. For trend data, comparative estimates for 91 countries from 1993 to 2013, 81 countries for 2008 to 2013, mostly in the developing world.

Source data: National household surveys and Luxembourg Income Study (LIS).

Description: Inequality estimates generated by the World Bank based mainly on household data, primarily income or expenditure household surveys (i.e. a largely microdata-based dataset, with grouped data used only for a few countries). The focus is mainly on developing countries. The inclusion of developed countries is recent. Almost all the poverty and inequality statistics reported in the WDI originate from PovcalNet (with the main exception being national poverty lines/rates). For developing countries, the World Bank uses consumption rather than income when possible. Income distribution and Gini indexes for high-income countries are calculated directly from the LIS database, using an estimation method consistent with that applied for developing countries. *No access to underlying microdata.*

Indicators: Gini index, annualized average growth rate in per capital mean income (or consumption) for bottom 40 per cent of population, annualized average growth rate in per capital mean income (or consumption) for total population, income share held by highest 10 per cent, income share held by lowest 10 per cent, income share held by lowest, second, third, fourth and highest 20 per cent, survey mean income (or consumption) per capita of bottom 40 per cent of population (2011 PPP \$ per day), survey mean income (or consumption) per capita of total population (2011 PPP \$ per day).

Interpolation: None

Format for download: Spreadsheets in Excel, .csv, .xml available for download through the World Bank Open Data website (<https://data.worldbank.org>, last accessed 24 September 2018).

II. Luxembourg Income Study Database (LIS)

Host institution: Cross-National Data Center in Luxembourg, hosted by the University of Luxembourg and The Graduate Center of the City University of New York

Coverage (number of countries and years): About 50 countries over 10 waves (1980, 1984, 1990, 1995, 2000, 2004, 2007, 2010, 2013, 2016, approximately). Data includes 33 high-income countries, 13 middle-income countries (Brazil, China, Colombia, Dominican Republic, Hungary, Mexico, Panama, Paraguay, Peru, Romania, Serbia, South Africa, Tunisia), and 4 lower-middle income countries (Egypt, Georgia, Guatemala, India).

Source data: National household surveys.



Description: The main databases consist of harmonized microdata at the individual and household levels presented in a common framework. The LIS also offers a list of “key figures” (in Excel) free of charge. **Overall, this is the only source that provides inequality statistics calculated using a uniform set of assumptions and definitions on the basis of thoroughly harmonized microdata to maximize comparability.** Financially contributing countries and contributing institutions (OECD, ILO, IMF and World Bank) are granted unlimited access to the microdata.

Indicators: LIS datasets contain multiple household- and person-level variables on labour income, capital income, social security and private transfers, taxes and contributions, demography, employment, and expenditures. The “key figures” table includes Gini and Atkinson coefficients (disposable income only), percentile ratios, relative poverty rates, distribution of children by income groups, child poverty.

Interpolation: None

Format for download: Aggregate indicators and key figures available in Excel for download from: <https://www.lisdatacenter.org/lis-ikf-webapp/app/search-ikf-figures>. For member countries and institutions, microdata can be downloaded from the LIS data centre website in SPSS, SAS, Stata and R.

III. Standardized World Income Inequality Database (SWIID, v. 7.1, August 2018)

Author: Frederick Solt (University of Iowa).

Coverage (number of countries and years): 192 countries for as many years as possible from 1960 to the present.

Source data: LIS, OECD Income Distribution Database, the Socio-Economic Database for Latin America and the Caribbean (CEDLAS and the World Bank), Eurostat, PovcalNet, UN ECLAC, national statistical offices around the world and academic studies.

Description: Datasets of Gini coefficients calculated through a multiple imputation method using both microdata-based sources (e.g. LIS) and secondary sources. Because of the imputation of missing country/year cells, SWIID has the largest coverage of any of the datasets described in this note. The LIS definitions of income are employed as the standard for the imputation procedure. Comparability across datasets is assessed in terms of success in predicting LIS data or estimates as well. The author minimizes reliance on problematic assumptions by using as much information as possible from proximate years within the same country for the imputation of missing values. However, he relies on the assumption that ratios between different inequality measures are constant, or stable, and can therefore be used to predict those variables when they are not observed. Stata code on the SWIID website allows users to either calculate means, standard deviations and confidence intervals, or to use the 100 values produced by multiple imputations directly. *No access to underlying microdata.*

Indicators: Gini coefficients of market and disposable income, absolute redistribution (market-income inequality minus net-income inequality), and relative redistribution (market-income



inequality minus net-income inequality, divided by market-income inequality), along with estimates of uncertainty in these statistics. SWIID net income is post-tax, post-transfer, and benchmarked to LIS Key Figures. SWIID market income is pre-tax, pre-transfer, and benchmarked to an income inequality series generated from LIS microdata.

Interpolation: Bayesian

Format for download: Online generator allows for comparisons of indicators, which can also be downloaded as a comma separated value file readable in Excel. Datasets formatted for use in Stata and R are available for download.

Downloads available from: <https://dataverse.harvard.edu/dataverse/fsolt> (last accessed on 24 September 2018).

IV. World Income Inequality Database (WIID, v. 3.4, January 2017)

Host institution: UNU-Wider.

Coverage (number of countries and years): 182 countries, for various years, beginning in the 1940s. Most data series start in the 1980s, however. Data are not available for many developing countries for extended periods of time.

Source data: household survey statistics obtained from national statistical offices, the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), the OECD Income Distribution database (IDD), the EU-Statistics on Income and Living Conditions (EU-SILC), LIS and PovCalNet.

Description: WIID is a secondary dataset that compiles country-year estimates from many databases. As such, it is mainly a repository of data from other sources, although it categorizes the inequality statistics by income concept and equivalence scale, and it provides an assessment of quality. The onus is on the users to determine which of the many estimates provided should be used in their research. The dataset comes with rich metadata and documentation files. *No access to underlying microdata.*

Indicators: Gini, deciles and quintiles of income/consumption.

Interpolation: None

Format for download: Available for download in Excel and Stata from: <https://www.wider.unu.edu/database/world-income-inequality-database-wiid34>. A new version of the data set should be available “quite soon”. Early access for DESA could possibly be arranged.



V. World Wealth and Income Inequality Database (WID.world)

Host institution: World Inequality Lab

Source data: Fiscal (income tax) data and data from national accounts combined with other sources (household income and wealth surveys, inheritance and wealth tax data, as well as wealth rankings published in the media, e.g. Forbes rankings).

Coverage (number of countries and years): 70 countries according to website (57 countries on own count), mostly 1980-2016, but with much older data (1920s) for some countries.

Description: Household surveys do not properly capture levels and trends in the income and wealth of the richest individuals. WID combines national accounts and fiscal data with survey data to overcome this limitation. The database provides annual estimates of the distribution of income and wealth (“distributional national accounts”) harmonized based on definitions of income and wealth that are consistent with the macroeconomic national accounts. Most attention goes to accurately estimating the income share of the top decile.

Indicators: Income and wealth distribution, including Gini, income and wealth shares of the top 1%, 10%, bottom 10%, 50%, middle 40%, ...; wealth by public-private ownership.

Interpolation: Pareto

Format for download: Available for download in Excel from <https://wid.world/data/>; Stata command for direct access from internet.

VI. All the Ginis (ATG, v. October 2016)

Author: Branko Milanovic (The Graduate Center at CUNY)

Source data: LIS for 46, mostly developed, countries; SEDLAC for 23 countries in Latin America; EU-SILC; World Bank’s Eastern Europe and Central Asia dataset; World Income Distribution (WYD) dataset; PovcalNet; WIID (using only the observations that are conceptually comparable to those in the other sources used by the author) and individual datasets for specific countries from research studies.

Coverage (number of countries and years): 166 countries; as many years as possible 1950-2015 (most series end in 2014).

Description: Gini coefficients from the sources listed (i.e. based on household surveys), standardized by the author and used to create a set of “preferred Ginis” based on a selected set. The underlying Gini series for the eight sources are also provided. The dataset includes information for each country-year on the welfare concept used (income or consumption; market/gross or net/disposable income) and recipient unit (individual or household). It also shows the source of information used for each country-year. This is the only income inequality dataset that uses published expert papers to complement the information provided by cross-



country databases. The dataset comes with a straight-forward description (in pdf). *No access to underlying microdata.* The dataset was last updated in October 2016.

Indicator: Gini coefficient.

Interpolation: None.

Format for download: Stata. From <https://www.gc.cuny.edu/Page-Elements/Academics-Research-Centers-Initiatives/Centers-and-Institutes/Stone-Center-on-Socio-Economic-Inequality/Core-Faculty,-Team,-and-Affiliated-LIS-Scholars/Branko-Milanovic/Datasets>.

VII. Estimated Household Income Inequalities (EHII)

Authors: University of Texas Inequality Project (UTIP) (research team led by James Galbraith)

Source data: UTIP-UNIDO global dataset on industrial pay inequality, World Bank's Deininger and Squire (1996) dataset.

Coverage (number of countries and years): 147 countries, 1963-2015.

Description: Gini and Theil indexes of gross/market income inequality. These are not measured directly using survey data or the other sources described in this note but use a UTIP-UNIDO dataset of industrial pay inequalities ("between-groups components of Theil's T statistics calculated across industrial categories") as the primary source of information. Specifically, the calculations are based on a regression that shows the relationship between inequalities of industrial pay and household income inequalities, as measured in the World Bank's DS dataset.³ The dataset does not come with a description of the methodology used to estimate income inequalities from industrial pay data.⁴ *No access to underlying microdata.*

Indicators: Gini coefficient, Theil index.

Interpolation: unclear.

Format for download: Excel spreadsheet available for download here: <https://utip.lbj.utexas.edu/data.html>

VIII. Global Consumption and Income Project (GCIP)

Authors: Arjun Jayadev (UMASS Boston), Rahul Lahoti (University of Goettingen), Sanjay G. Reddy (New School for Social Research).

Source data: The project builds on various existing data resources, including the LIS.

³ The World Bank's Deininger and Squire dataset was produced in 1996. It is not clear whether the authors are referring to updated inequality estimates using the World Bank's WDI or rely on old (1996) estimates.

⁴ The following article gives some information: Galbraith, James K. and others (2015). A comparison of major world inequality data sets: LIS, OECD, SILC, WDI and EHII. UTIP Working Paper no. 69, March 2015. However, it does not describe the methodology used to produce Gini coefficients of income inequality.



Coverage (number of countries and years): unknown.

Indicators: unknown.

Interpolation: unknown.

Format for download: Currently unavailable (as of October 2018).

Differences across datasets—some examples

In general, different databases show similar trends in inequality for most countries.⁵ But there are instances where levels of inequality in particular countries differ across databases. Data and estimates for developed countries are relatively consistent. Discrepancies are particularly high in estimates for sub-Saharan Africa and, to a lesser extent, for Asia.

Figure 1 compares Gini coefficients of market/pre-tax income for Brazil, China and the United States using SWIID and World.WID. In China, according to SWIID, the Gini of market income increased from 41.7 in 2001 to 46.0 in 2015. During the same period, World.WID marks an increase from 49.6 to 54.8. Both databases register a decline in inequality in China but differ substantially in the actual value of the Gini coefficients in each year. Both databases show Brazil as the most unequal country of the three, with the highest Gini coefficient, although the values themselves differ (53.7 in the SWIID versus 62.5 in World.WID).

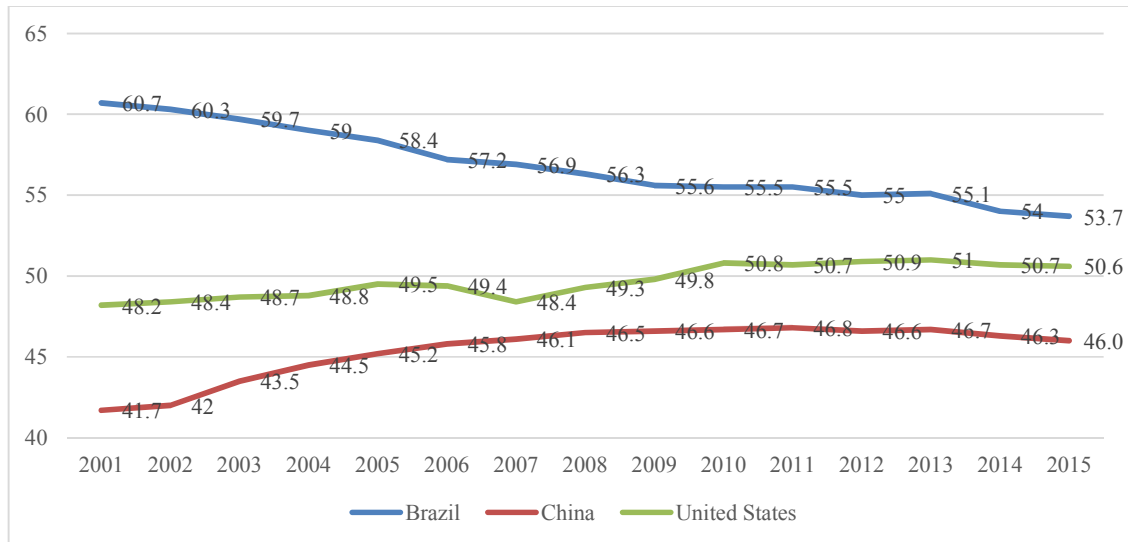
Figure 2 compares Gini coefficients of disposable income using SWIID and PovcalNet. Levels of inequality are higher according to PovcalNet, especially in Brazil and in the United States (Ginis for both sources rely on income data).

⁵ Ferreira, Francisco, Nora Lustig and Daniel Teles (2015). “Appraising cross-national income inequality databases: an introduction”, *Journal of Economic Inequality*, Vol. 13, issue 4, Dec. 2015.



Figure 1. Trends in Gini coefficients of market income in Brazil, China and United States, 2004-2015, using SWIID and World.WID.

A. SWIID



B. World.WID

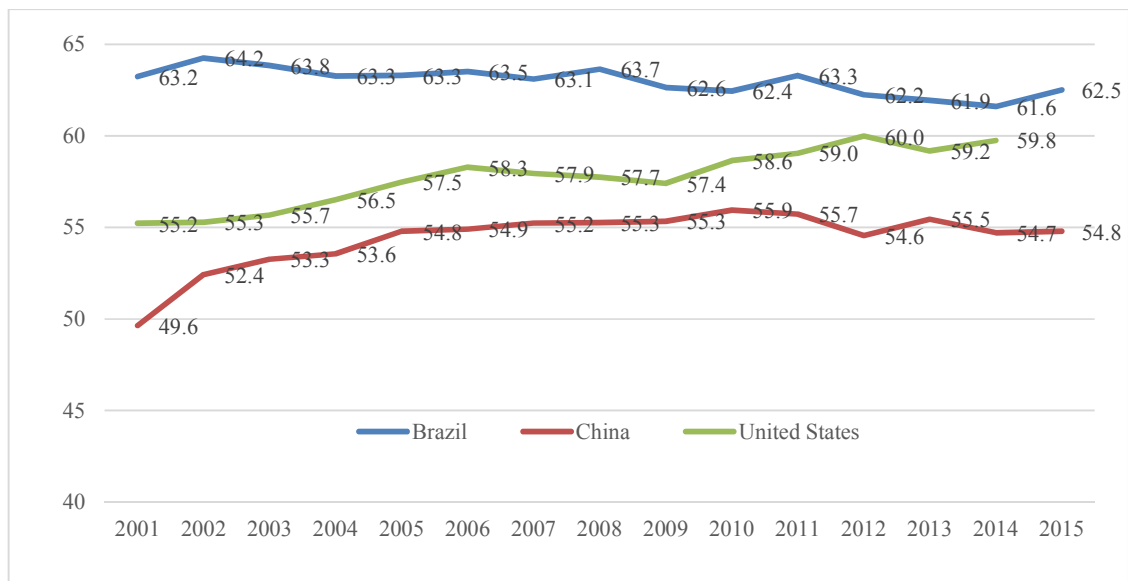
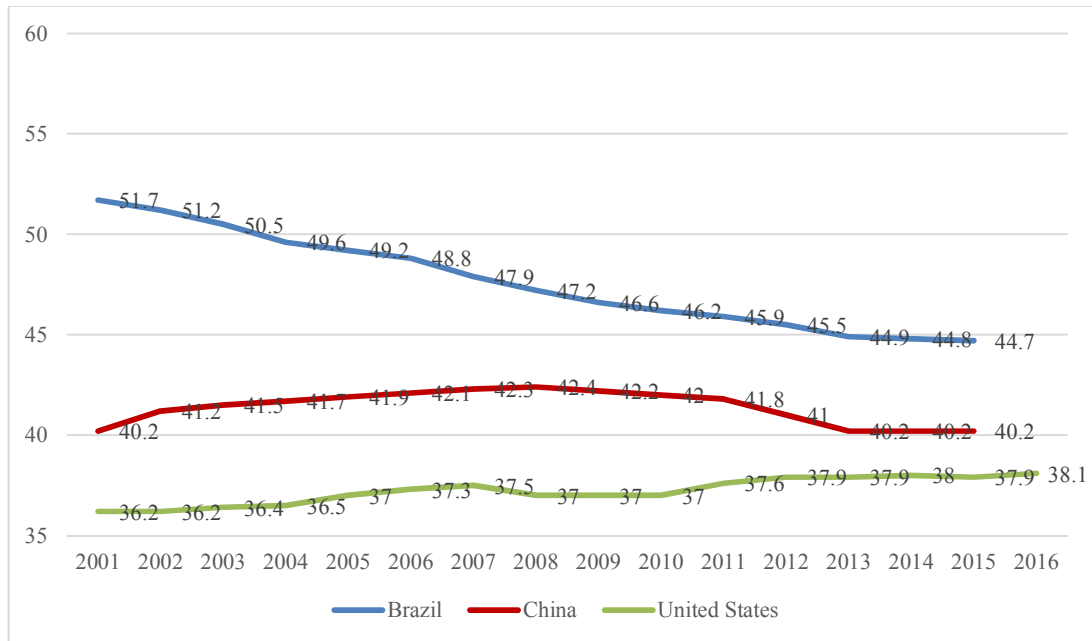


Figure 2. Trends in Gini coefficients of disposable income in Brazil, China and United States, 2004-2015, using SWIID and PovcalNet

A. SWIID



B. PovcalNet

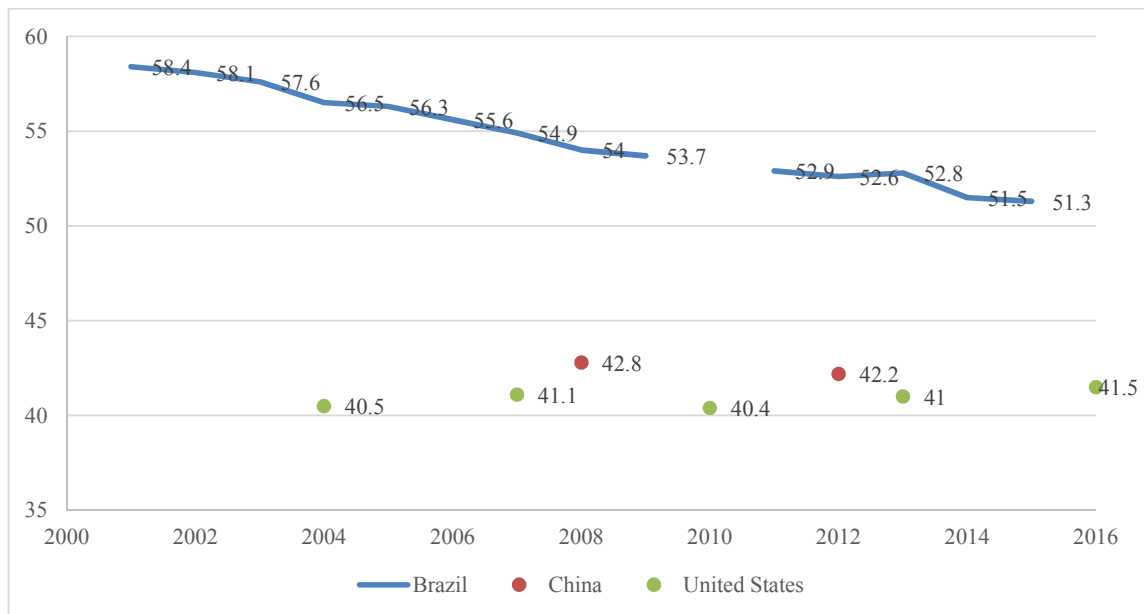


Figure 3 also shows significant average differences in Gini coefficients estimated using different data sources. Figure 4 indicates that most sources agree regarding the direction of changes in inequality, but trends differ in a few cases.

Figure 3. Robustness check: differences in Gini coefficients at six points in time, taking PovcalNet as reference point

Source		1988	1993	1998	2003	2008	2013
LIS	Average diff	3.03	2.77	2.57	2.53	2.97	3.01
	N	19	24	27	25	10	12
SWIID	Average diff	2.26	1.86	1.8	1.24	1.41	2.21
	N	72	101	105	129	120	65
OECD	Average diff	2.23	1.16	2.21	2.05	1.79	1.81
	N	6	11	11	23	29	18

Source: Ferreira, Francisco H.G. and others. Inequality increasing everywhere? Evidence from a global database of household surveys. Presentation given at the United Nations expert group meeting, *New research on inequality and its impacts*, 12-13 September 2018, New York. Available from: <https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2018/09/Ferreira.pdf>.

Figure 4. Robustness check: country trends in Gini coefficients from 1993 to 2008

Database	↑↑	↑	?	↓	↓↓	Disagreement	Total
PovcalNet	42		10		39		91
PovcalNet vs. Eurostat	1	2	2	1	1	0	7
PovcalNet vs. LIS	4	1	0	0	2	0	7
PovcalNet vs. SEDLAC	6	0	0	0	6	0	12
PovcalNet vs. CEPAL	1	0	0	0	4	2	7
PovcalNet vs. SWIID	34	8	5	9	27	5	88
PovcalNet vs. (Eurostat + LIS + SEDLAC)	11	3	2	1	8	0	25

- ↑↑ both PovcalNet & alt. source show rise greater than 1 point
- ↑ one database shows rise of more than 1 point, while the other has change within 1 point
- ? Gini changed by less than 1 point in both
- Disagreement They go in opposite directions (change exceeds 1 point)

Source: Ferreira, Francisco H.G. and others. Inequality increasing everywhere? Evidence from a global database of household surveys. Presentation given at the United Nations expert group meeting, *New research on inequality and its impacts*, 12-13 September 2018, New York. Available from: <https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2018/09/Ferreira.pdf>.

