J Econ Inequal (2015) 13:629–671
DOI 10.1007/s10888-015-9305-3

World income inequality databases: an assessment of WIID and SWIID

Stephen P. Jenkins[1,2,3]

Received: 27 May 2015 / Accepted: 29 May 2015 / Published online: 22 August 2015

Abstract This article assesses two secondary data compilations about income inequality – the World Income Inequality Database (WIIDv2c), and the Standardized World Income Inequality Database (SWIIDv4.0) which is based on WIID but with all observations multiply-imputed. WIID and SWIID are convenient and accessible sources for researchers seeking cross-national data with global coverage for relatively long time periods. Against these undoubted benefits must be set costs arising from lack of data comparability and quality and also, in the case of SWIID, questions about its imputation model. WIID and SWIID users need to recognize this benefit-cost trade-off and ensure their substantive conclusions are robust to potential data problems. I provide detailed description of the nature and contents of both sources plus illustrative regression analysis. From a data issues perspective, I recommend WIID over SWIID, though my support for use of WIID is conditional.

Keywords Global inequality · Inequality · Gini · Imputation · WIID · SWIID

1 Introduction

This article assesses two 'World Income Inequality' databases: WIID (version 2c, May 2008) produced by UNU-WIDER (2008), and the 'Standardized WIID' (SWIID, version 4.0, September 2013) produced by Frederick Solt (2013a) which is based on WIID supplemented by other sources but is distinctive for having all of its observations multiply-imputed. Both WIID and SWIID are secondary data sets that compile country-year

✉ Stephen P. Jenkins
s.jenkins@lse.ac.uk

1 London School of Economics, London, UK

2 University of Essex, Colchester, UK

3 IZA, Bonn, Germany

estimates of summary measures of income distributions (inequality summarized by the Gini coefficient in particular). WIID and SWIID are notable in terms of their coverage in terms of numbers of countries (161 in WIID, 173 in SWIID) and years (from 1867 to 2006 in WIID; 1980 to 2012 in SWIID). For researchers seeking cross-national data with global coverage for relatively long time periods, WIID and SWIID are convenient and accessible sources. Against these undoubted benefits must be set costs arising from lack of data comparability and data quality more generally.

This article illustrates these data issues in order to bring them to the attention of current and potential users, taking the potential benefits for granted. I argue that researchers employing WIID and SWIID data need to recognize the benefit-cost trade-off and to ensure that any substantive analytical conclusions that they draw are robust to data issues. I provide detailed description of the nature and contents of both sources plus illustrative analysis, benchmarking them against other sources where possible. This leads me to recommend WIID over SWIID from a data issues perspective, but my support for use of WIID is conditional in ways that I spell out later.

A comprehensive review of a predecessor of WIID – the Deninger and Squire (1996) dataset – and a more general discussion of the 'promise and pitfalls' of secondary data sets on inequality has already been provided by Atkinson and Brandolini (2001, 2009). My assessment of the current version of WIID inevitably follows in Atkinson and Brandolini's footsteps. I revisit the issues that they raise and argue that their cautionary conclusions still apply. Since SWIID is derived from WIID, many of the same conclusions also apply to that source.

There are also new issues to be addressed. SWIID has the feature of 'filling in the gaps' using a multiple imputation procedure. Any costs arising from its implementation need to be taken into account alongside the potential benefits arising from the greater coverage. The value of SWIID is contingent on the plausibility of the assumptions underlying the imputation model (issues of potential bias, broadly speaking) and proper use of the multiply-imputed data (issues of precision). I shall argue that questions can be raised about the imputation model in particular.

WIID and SWIID are used by social scientists from a range of disciplines, and are widely known about and accessible. My initial web search on 'summary inequality databases' led to around 22,300,000 results with the 'UNU-WIDER download' page for WIID listed first and the 'Standardized World Income Inequality Database' home page listed third (Google search, 31 January 2014). My search on 'WIID' led to about 1,380,000 results and straight to the 'UNU-WIDER: Database (WIID)' page; searching on 'SWIID' led to about 14,200 results and straight to the 'The SWIID - MyWeb' page (Google search, 31 January 2014).

There are three main types of study using these secondary data on income distribution, with the first two being the most common. The first includes analysis of the global distribution of income, that is inequality (or some other feature) of the income distribution at the global level, including trends over time, and differences within or between regions. Examples include Sala-i-Martin (2006) who examined convergence in the distribution of world income using non-parametric density estimation methods applied to WIID data about quintile group income shares. A more recent study in the same spirit but using parametric models is by Chotikapanich et al. (2012). Convergence of the global income distribution is also analysed by Clark (2013) but using SWIID. Gruen and Klasen (2008, 2012) study trends in inequality-adjusted measures of social welfare using WIID. For references to earlier studies using cross-national inequality databases, see Atkinson and Brandolini (2001, 2009).

The second main type of study involves econometric analysis of country panels in which a measure of inequality is used as the outcome variable to be modelled or, more commonly, as a variable explaining some other outcome. An example of the first type of study is by Teulings and van Rens (2008) who relate inequality to schooling returns using WIID. Another example is by Acemoglu et al. (2015) who examine the impact of democracy on inequality using SWIID.

Many of the second type of studies consider whether higher inequality is associated with more or less economic growth. Much-cited papers by Barro (2000) and Forbes (2000) examined this question using the Deninger and Squire (1996) data. Barro (2008) revisited the topic using an early version of the WIID and later versions have been employed more recently by Berg et al. (2012), Castelló-Climent (2010), and Chambers and Krause (2010). A February 2014 study by IMF researchers (Ostry et al. 2014), finding that lower inequality was correlated with faster growth, and which received much media publicity, drew on SWIID for its inequality data. In this paper, I consider the relationship between inequality and inflation and unemployment in my regression case study.

The third and less common type of study is based on individual-level data from cross-nationally harmonised cross-sectional surveys (such as the World Values Survey) in which the data from the various countries (and possibly multiple survey rounds) are pooled, and some individual-level outcome is modelled using both individual-level and country-level explanatory variables. Economic inequality is an example of the latter. I am aware of no WIID-based study taking this approach, but see Layte's (2012) study of the relationship between individuals' mental health and inequality using European data. SWIID is used as the source of inequality data in Solt's (2011) analysis of the relationship between individuals' nationalist sentiments and their country's economic inequality.

In Section 2 I reprise the principal issues raised by Atkinson and Brandolini (2001) and in the rest of the paper I show that they are still relevant. Section 3 is devoted to WIID and Section 4 to SWIID. In each case, I describe the database and documentation, coverage and content, and provide evaluative commentary. In Section 5, I discuss illustrative regression analyses using both WIID and SWIID in order to highlight issues raised in the earlier sections. My conclusions appear in Section 6. Like Atkinson and Brandolini (2001, 2009), much of my discussion is illustrated using data for rich countries, especially OECD and EU ones, because alternative inequality series are readily available with which to benchmark WIID and SWIID, and because I am most familiar with these countries' income distributions. However, I discuss data for developing countries at several points throughout the paper.

The way in which I explore and discuss WIID and SWIID is influenced by the fact that I had never used either of them before embarking on this paper. What I describe is the experience of a new user discovering what is in the data rather than a critique of substantive analyses that have been done with them. The commentary is forensic and specific on occasion but an important part of my message is that The Devil is in the Detail.

2 Data comparability issues raised by Atkinson and Brandolini

Atkinson and Brandolini (2001) highlight issues of data comparability. These are closely related to issues of data quality (which they also discuss in detail) since differences in quality across country-year observations lead to non-comparability. More generally, non-comparabilities may arise because of differences in the *definitions* of the ‘income distribution’, and also because of differences in the *data sources* and in the *processing* of the income data in the source. There may be differences in the series not only between countries in any year, but also changes over time for a given country. The combination of different definitions, the nature of the data sources, and their processing leads to what Atkinson and Brandolini describe as a ‘bewildering variety of estimates’ (2001: 784), which makes the selection of database observations a complex task for any user. I elaborate and summarize their points in the rest of this section in order to provide a reference point for my assessments of WIID and SWIID.

2.1 Differences in the definition of the ‘income’ distribution

The definition of the ‘income’ distribution involves variations along five main dimensions. First, there is the *resource definition*. The principal alternatives here are ‘income’ and ‘consumption’ (consumption expenditure). There is no decisive case in favour of one measure or the other: there are arguments to be made for both in terms of principle and of data collection. In practice, income measures are more commonly available for high-income countries, and expenditure measures for low-income countries. Regardless of which resource measure is chosen, there are potential differences in what might be included in the measure and questions about the comprehensiveness of coverage. For example, for income, major differences concern the treatment of personal income taxes (national or local) and related deductions such as employee social insurance contributions and of cash benefits (‘transfers’) received from the government. Market (or ‘original’) income includes none of these sources; pre-tax post-transfer (‘gross’) income includes cash benefits but does not deduct tax payments; post-tax post-transfer (‘disposable’ or ‘net’) income includes both. To give a concrete example, official distribution statistics in European countries are typically based on a disposable income definition, whereas the US Bureau of the Census uses a gross income definition. There are similar issues regarding the comprehensiveness of any consumption expenditure measure, including for example the treatment of spending on durables.

Second, there is the *reference period*, the time period to which the measure of income or consumption refers. For example, spending data derived from diary data often refer to spending over a period of less than one month. Income data may refer to the most recent pay period (as in UK surveys about earnings, and may be as short as a week or fortnight), or to the month or the year (‘annual income’). Third, there is the *reference unit*. Income and consumption can potentially refer to aggregates at the level of the household, the nuclear family, the tax unit, or indeed the individual. Fourth, there is the issue of adjustment for differences in reference unit size and composition (‘*equivalization*’). Income measures are often deflated by an equivalence scale to account for the fact that $5000 per month (say) provides higher living standards to a single person than to a family of four. Adjustments in practice range from no adjustment at all through to a per capita adjustment with many variations in between. An equivalence scale commonly used in Europe nowadays is the modified-OECD one, equal to one for the first adult in the reference unit, 0.5 for each additional adult, and 0.3 for each dependent child.

Fifth, there is the *unit of analysis*. The issue is whether each reference unit receives a weight of one or a weight equal to the number of individuals within the unit when the distributional summary statistics are derived. Compare, for instance, the distinction between the inequality of the distribution of household income among households and the inequality of the distribution of household income among individuals (each individual is assumed to receive the income of the household to which he or she belongs).

2.2 Differences in data sources

With regard to differences in data sources, Atkinson and Brandolini (2001) point to aspects of intrinsic data quality and reliability. These include issues of population coverage (all individuals within a country versus only urban households, or individuals with incomes above the income tax threshold, for instance) and representativeness, non-response, and measurement error. There may be different types of data sources (e.g. surveys or tax administration records), and there may be multiple sources available for a given country-year observation as well.

2.3 Differences due to data processing

Atkinson and Brandolini (2001) draw attention to the fact that a given data source may be used in a variety of ways to derive income distribution statistics. Calculations may be made from unit record data or from published tabulations (banded data). In the former case, there may be different versions of the same source utilized, as for example in the USA where the Bureau of the Census calculates distributional statistics using an 'internal' (more detailed) version of the Current Population Survey, whereas only less-detailed 'public use' data are readily available to most researchers. Income data may be top coded in the source (values greater than a particular threshold set equal to the threshold value) and different assumptions may be made about how to handle extreme values, for example the treatment of units with zero or negative recorded incomes, or high-income outliers. These are issues of censoring (right and left) and truncation ('trimming'). In the case of banded data, potential differences may arise if there are changes over time in the numbers of income groups and the group boundaries, and from differences in the methods used to estimate summary income distribution statistics from the published data.

2.4 The implications of the differences

Many of these data differences have predictable effects on inequality. Other things being equal, one would expect the inequality of consumption to be less than the inequality of income, the inequality of disposable income to be less than the inequality of market or gross income (reflecting the redistributive nature of taxes and transfers), and inequality among households to be lower than inequality among nuclear families, and inequality to be lower the longer that the reference period is. (Varying the equivalence scale has ambiguous effects on inequality, however: see Coulter et al. 1992.) Trimming data for outliers is likely to reduce inequality; making imputations for right-censored (top coded) observations will increase inequality.

The problem is that other things are not equal in secondary data set compilations: there is substantial heterogeneity across countries and across years and the researcher has only the secondary data to hand rather than the original sources. Nevertheless, the various data issues are of little consequence if they have little impact in practice – but arguably they do. In this paper, I use a difference of two percentage points between a pair of Gini coefficient estimates as a signal of a difference that needs to be investigated. This benchmark is chosen because year-on-year changes in a country's Gini coefficient are only rarely this large.

Atkinson and Brandolini (2001) show that the preferred ('accept') series in the Deninger and Squire (1996) data set leads to different conclusions about cross-national inequality rankings among OECD countries at a point in time, and different conclusions about within-country trends in inequality over time, than are produced by other series of at least as good a quality. The relationship between inequality (measured by the Gini coefficient) and price inflation is also shown to be sensitive to choice of inequality data series that is used. The non-robustness theme is illustrated at greater length by Atkinson and Brandolini (2009) with, *inter alia*, extended analysis of the relationship between income inequality and globalization estimated using regression analysis of time series data for a panel of 16 OECD countries (an example of the second type of study identified in the Introduction).

In the light of these issues of data quality and comparability, Atkinson and Brandolini (2001) make recommendations about both the *construction and development* of secondary data sets on income distribution, and their *use*. Under the first heading, the emphasis is on provision of full documentation of sources for each series and construction of any derived variables, together with additional variables enabling users classify estimates according to the headings identified above. Multiple observations for each country-year need to be justified in terms of value-added, and redundancies eliminated. The emphasis on data consistency and understanding of national data sources is re-emphasized by Atkinson and Brandolini (2009), who suggest that 'this may lead us to analyse a carefully matched subset of countries, rather than to seek to maximize their number' (2009: 400).

Under the second heading, Atkinson and Brandolini (2001) discuss the commonly-used 'dummy variable adjustment' method for handling data differences in regression analysis. This is where country-year data employing multiple income definitions are pooled but dummy variables are used to identify observations based on definitions other than the reference one. (Alternatively, researchers run first-stage regressions to standardize for definitional differences in the inequality measure, and use the standardized predictions of it in the main analysis.) For data observations based on gross and net income, for example, the procedure effectively assumes that the absolute difference between inequality measured using one income concept and inequality measured using another concept is constant across time and across countries: there are simple intercept shifts. This is implausible because the extent of redistribution – commonly measured by such a difference – varies across countries and time (OECD 2011: Chapter 7).

Atkinson and Brandolini (2009) discuss the adjustment method more generally using detailed illustrations, and caution against its mechanical application, recommending instead 'using a data-set where the observations are as fully consistent as possible' (2001: 790). This approach to sensitivity analysis is illustrated by them (see their Appendix) and is also taken recently by, for example, Castelló-Climent (2010). The approach may be contrasted with the dummy variable adjustments by Gruen and Klasen (2008, 2012) and Teulings and van Rens (2008), or the manual adjustments to the same effect by Chambers and Krause (2010). I discuss such adjustments further below.

Against this background, I now turn to assess the extent to which the issues raised by Atkinson and Brandolini with reference to the Deninger and Squire (1996) data set and earlier versions of WIID remain relevant.

3 The World Income Inequality Database (WIID2c)

The best short introduction to WIID is the description on its home page UNU-WIDER (2008):

World Income Inequality Database V2.0c May 2008

The UNU-WIDER World Income Inequality Database (WIID) collects and stores information on income inequality for developed, developing, and transition countries. The database and its documentation are available on this website.

WIID2 consists of a checked and corrected WIID1, a new update of the Deininger & Squire database from the World Bank, new estimates from the Luxembourg Income Study and Transmonee, and other new sources as they have became available. WIID2a contains fewer points of data than WIID1 as some overlaps between the old Deininger & Squire data and estimates included by WIDER have been eliminated along with some low quality estimates adding no information. In addition to the Gini coefficient and quintile and decile shares, survey means and medians along with the income shares of the richest 5% and the poorest 5% have been included in the update. In addition to the Gini coefficient reported by the source, a Gini coefficient calculated using a new method developed by Tony Shorrocks and Guang Hua Wan is reported. The method estimates the Gini coefficient from decile data almost as accurately as if unit record data were used.

Source: http://www.wider.unu.edu/research/Database/en_GB/database/ with emphasis in original. (Accessed 30 March 2014.)

A menu on the side of the webpage provides access to pages for Download (of the data), Income distribution links (to the Luxembourg Income Study, Transmonee, and SEDLAC), Frequently Asked Questions, WIID documentation, and Country documentation. WIID documentation consists of a 44-page downloadable pdf file 'giving a general description of the database and its contents' (20 pages of which contain References), plus two files with brief 'Revision notes of latest updates' (they summarize the changes from versions 2a through to the current 2c). The Country documentation is a series of documents that 'provide information about the sources and the surveys used as far as documentation was available', downloadable in pdf format. A drop down menu accesses the sheet for each country. Each has to be read or downloaded separately and some sheets appear to be unavailable. (I did no systematic checks but two sheets that I found unavailable on 30 March 2014 were those for the United Kingdom and Vietnam).

3.1 WIID: data and documentation

The WIID data are in a 1.76MB Excel spreadsheet. Eager to check whether I could simply 'plug and play' with the data, I imported them into Stata version 13.1 with the command `import excel, firstrow`, and then checked the variables available and their characteristics.

Much was as expected: there were *Country* and *Year* variables, other variables with names apparently corresponding to the income distribution statistics cited in the home page blurb cited above, together with variables identifying definitional differences (there are variables with names corresponding to each of the five headings identified in the previous section: *IncDefn*, *Curref*, *IncSharu*,*Equivsc*, *UofAnala*) and variables with names referring to sources (e.g. *Source1* and *SurveySource2*), and dimensions of coverage (*AreaCovr*, *PopCovr*, *AgeCovr*). There were also variables suggestively labelled *Quality* and *Revision*. A listing showed that the country-year observations were ordered alphabetically by country but not by year within-country. There were 5,313 country-year observations, for 161 distinct countries and 88 distinct years.

Since I have analysed UK inequality data extensively (using mostly national sources), I was keen to see what was in WIID for the UK. The 99 *ReportedGini* estimates are shown, by year, in Fig. 1. It was immediately clear that most of the UK estimates refer to the period after 1960, which was not surprising given my knowledge about the data sources available. Perhaps more surprising – despite my reading of Atkinson and Brandolini (2001) – was the prevalence of multiple observations per year and the wide range spanned by the estimates at these points, even if one distinguished between observations of $Quality = 1$ ($N = 70$) and the rest ($Quality = 2,3,4$; $N = 29$). I rapidly decided that attempts at 'plug and play' with WIID are pointless. Reading the documentation is essential to distinguish the data points and to undertake any analysis. In particular, I needed to confirm whether $Quality = 1$ was the highest quality classification (as I guessed) or the lowest (it is the highest).

Even this brief exploration suggests some ways to improve the usability of WIID in a later release. Although the spreadsheet data format used to provide the data is portable, it is restrictive and prohibits even cursory documentation being associated with the data series. Variable names are generally sufficiently evocative of content, but it would be better to supplement names with meaningful variable labels. Variables such as the *Country* identifier and those defining the data could be converted from text to numeric, and the existing text used as the label, thereby also saving storage space. This would also be a good opportunity to identify missing values consistently. I would prefer *Curref* information about reference period and currency unit to be in two variables, not one – they are distinct concepts.

The content of text (string) variables should be proof read and inconsistencies in spelling removed. Misspellings in variables can lead to different series being identified by mistake. (In what follows, I use data which I corrected for some obvious typographical inconsistencies.) Categorical variables, including *Quality* or *Version*, need value labels. Variable and value labels can be easily stored along with the data, were widely-used statistical software

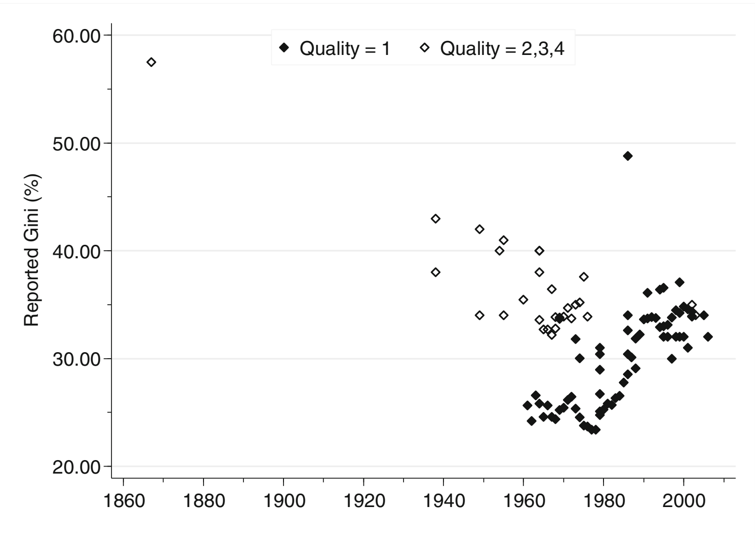


Fig. 1 WIID's 99 'Reported Gini' estimates for the United Kingdom. *Notes* The WIID four-category quality assessment variable (*Quality*) is explained in the main text. $Quality = 1$ is the highest quality category

such as SPSS or Stata to be used, and portability would not be lost because it is easy to swap between data formats nowadays. One variable (*AK*) can be deleted altogether: it has missing values for all observations. Variable display formats can be tidied up: for example, the *ReportedGini* (in per cent) includes two redundant decimal places (as shown in Fig. 1). Surprisingly, the crucial year identifier (*Year*) contains text rather than numeric content, and this turns out to arise because 22 country-year observations refer to multiple years, e.g. '1953–55' (all but two observations are for India; only one observation refers a period after 1980). I would recommend that such dates be converted to numeric values (e.g. midpoints of the period spanned), and labels or, better, a new flag variable be used to identify such cases. After all, similar decisions must have already been taken for financial years that span calendar years.

WIID's producers helpfully include three-letter country code identifiers (*Country3*) in the data along with full text country names. But they could go further to help users. The cross-national element is so fundamental that it would be useful to provide more information about the countries. I would also like to see two-letter country code identifiers (more useful for labelling in graphs, and it would assist merges with other databases using this code as an identifier), and variables that classify countries by geographic region and politico-economic memberships such as of the EU and the OECD (and dates of joining), etc. Data points could also be classified by period, again for user convenience. All these variables I ended up creating myself. It would help to also have the Country documentation available as a single pdf file.

After studying the documentation cited above and some further explorations of the data (described shortly), I believe that WIID has successfully implemented Atkinson and Brandolini's (2001) recommendations regarding *construction and development* of secondary databases (summarized in the previous section). Although Atkinson and Brandolini sought updates to databases that are documented accumulations of previous versions, the WIID producers persuade me that their approach which both adds new information and deletes only 'overlapping estimates and those that add no information' (WIID Documentation: 10) is satisfactory. Atkinson and Brandolini's (2001: 795) recommendation that estimates be classified to give users 'maximum guidance' has also been addressed by a revision of the data quality rating. This now has four categories with the highest (*Quality* = 1) referring to observations for which the underlying concepts are known and where the quality of the income concept and the survey can be judged 'sufficient' (WIID Documentation: 15). *Quality* = 2 for observations 'where the quality of *either* the income concept *or* the survey is problematic or unknown or [WIID] have not been able to verify the estimates' (*ibid*). *Quality* = 3 is the case when 'both the income concept and the survey are problematic or unknown' and *Quality* = 4 for memorandum items, 'given this rating since the data lying behind the observations often are unreliable' (*ibid*). In addition, 'some final guidelines' remind users to pay attention to definitional differences and not to simply combine observations of different types unless corrections or adjustments are employed, to consult the country sheets. Can users follow these guidelines with the resources provided?

Overall, my verdict is affirmative because the WIID documentation is reasonably comprehensive in explaining the variables included on the file. To be sure, tracking down the precise origins of the estimates included (the method of derivation, the publication that derived the estimate or the original data source used in the calculation) requires some detective work and time. Closer integration of the country documentation with the variable documentation, including the suggestions made above, would help in this direction.

3.2 WIID: coverage and content

I now turn to issues related to the use of WIID, beginning with its coverage of countries and years. See Table 1 for a summary. The top panel shows all of the observations by region and year; the bottom panel shows the same classification but only for high quality observations (*Quality* = 1). The main lessons are, first, the majority of observations refer to years after 1980 and to rich countries (Western Europe – defined here as the EU15 – and North America, which means Canada and the USA). Second, the number of *Quality* = 1 observations is only 26 % of the total. Third, selecting *Quality* = 1 observations further weights coverage towards rich western nations. There is a marked loss of observations from Africa and Central and South America and Asia, in particular. Researchers attempting global analysis using WIID face an uncomfortable coverage-quality trade-off that cannot be avoided when selecting countries and years.

In what follows, I restrict attention to *Quality* = 1 observations (following Atkinson and Brandolini who confined their analysis to the Deininger-Squire 'accept' data), with some

Table 1 WIID: number of country-year observations, by geographical region and year

Region	Period							
	1867–1899	1900–1959	1960–1969	1970–1979	1980–1989	1990–1999	2000–2006	*Total*
All observations								
Africa	0	28	61	56	67	140	26	*378*
Western Europe (EU15)	1	54	98	141	235	342	182	*1,053*
Other Europe, Turkey, Russia	0	11	68	72	185	483	231	*1,050*
North America	0	17	25	35	53	51	10	*191*
Central & South America	0	34	154	177	197	424	124	*1,110*
Central, East, & South East Asia	1	96	188	210	280	288	85	*1,148*
Oceania	0	42	42	43	45	55	11	*238*
Middle East	0	20	19	30	22	23	9	*123*
Total	*2*	*302*	*655*	*764*	*1,084*	*1,806*	*678*	*5,291*
Observations with Quality =1								
Africa	0	0	0	0	3	2	0	*5*
Western Europe (EU15)	0	2	19	72	163	293	170	*719*
Other Europe, Turkey, Russia	0	4	5	10	17	135	95	*266*
North America	0	14	16	28	44	42	9	*153*
Central & South America	0	0	0	2	15	40	8	*65*
Central, East, & South East Asia	0	0	5	15	39	53	8	*120*
Oceania	0	0	0	0	18	28	7	*53*
Middle East	0	0	0	2	2	13	3	*20*
Total	*0*	*20*	*45*	*129*	*301*	*606*	*300*	*1,401*

The classification excludes 22 country-year observations with multi-year 'year' values. All observations classified in the table have non-missing observations on Reported Gini. 'Quality = 1' refers to the highest WIID data quality classification. See main text for details exceptions when I discuss developing countries. I also restrict attention to inequality estimates derived for data sources with national geographical coverage and of all ages (variables *AreaCvr* = *AgeCvr* = 'All'). This reduces the total number of observations to 1,273 but the region-year coverage pattern is similar to that shown in the bottom panel of Table 1. It is this subset of observations that I refer to for brevity as 'high quality' in what follows.

What about the coverage of the various income distribution statistics? The most comprehensive set refers to Gini coefficients (non-missing for all 1,273 high quality observations). However, among this subset, there are only 143 estimates of the share of poorest fifth (Q5), 673 estimates of the share of the poorest tenth (D10), 674 means, and 585 medians. There are only 86 estimates of the income share of the richest 5% (mostly from North America) and only 50 of the income share of the poorest 5% (virtually all from the EU15). So, although the WIID team has put considerable effort in adding more types of distributional summary statistic, the number of observations added is relatively small if one restricts attention to high quality data.

I focus attention on Gini coefficients in what follows, and the Reported Gini (*ReportedGini*) in particular. WIID2c also provides estimates of *Gini*, the Gini coefficient which is the estimate calculated by the Shorrocks and Wan (2008) method for income share data if share data are available but which is set equal to the Reported Gini otherwise. The Reported Gini is the Gini reported by the original source or, if none were reported, the estimate calculated using the World Bank's POVCAL package by the WIID team or Deininger and Squire. The Pearson correlation between the two Ginis for the high quality observations is greater than 0.99, which is not surprising given the relatively small number of income share observations (so many refer to the same value by construction). Thus the value-added from inclusion of the *Gini* variable in the new version of WIID is relatively small.

Let us now consider the heterogeneity of definitions and data sources underlying the estimates. If observations are classified along the five dimensions relating to definition and the two relating to source (publication and data source) then, among the subset of high quality observations, there are 357 distinct series. This number exaggerates the degree of heterogeneity since country-specific observations often rely on national sources. However, if the classification is re-done using only the five income definition dimensions, there are still 137 distinct series. There are 11 for the UK, 13 for Finland, and 11 for the USA. Across the countries, there are more than 40 series for 1994 and 1995. Arguably, the impact of differences in the reference period and the currency unit (*Curref*) are less important than the other four dimensions when one is looking at relative inequality. Dropping this dimension, one gets 56 distinct series in total, but the number for the countries changes little: there are 9 series for the UK, 11 for Finland, and 10 for the USA. There are around 20 series for each of 1994 and 1995.

More evidence about heterogeneity is provided by the number of observations per country-year cell. These are summarized in Table 2 for the subset of high quality observations. In around 70 % of cells, there are least two observations, and about one fifth have five or more observations. The prevalence of multiple observations is greatest in the 1990s. The maximum number of observation per country-year is 10, occurring in two cells (Spain in 1973 and 1990).

The existence of multiple series and multiple observations per country-year is not a criticism of WIID as a data provider. Because one researcher may be interested in one distributional concept and another researcher in a different one, it is useful to have series that serve both users. WIID's responsibility is to clearly document the different series available (though if multiple observations for the same series for a given country-year are provided, perhaps only one should be retained).

Table 2 WIID: number of observations per country-year cell, by period

Number	Period						
	1900–1959	1960–1969	1970–1979	1980–1989	1990–1999	2000–2006	*Total*
1	18	25	46	63	112	139	*403*
2	0	8	24	60	140	78	*310*
3	0	3	6	63	87	27	*186*
4	0	0	12	20	80	24	*136*
5	0	0	5	30	70	15	*120*
6	0	0	6	6	48	0	*60*
7	0	0	0	0	7	7	*14*
8	0	0	0	16	8	0	*24*
10	0	0	10	0	10	0	*20*
Total	*18*	*36*	*109*	*258*	*562*	*290*	*1,273*

High quality observations only. There can be more than one observation per countryyear cell because WIID contains multiple series for each country. See main text for details

For researchers, the implication is that careful selection of WIID observations is required for any sort of analysis, paying close attention to the different definitions and sources. This lesson is illustrated by the case of Finland, a country with multiple data series and a high prevalence of multiple observations per country cell. Figure 2 displays trends between 1960 and 2006 of the Reported Gini for the 13 data series characterised by the five WIID variables. (See the notes to the figure for further explanations.) For six series, there is only one observation, and two series provide four observations. It turns out that three of the four longest series, spanning 1986–2003, all derive from one source. This is 'Statistics Finland 2005' according to *Source1*, which turns out to be a webpage in Finnish cited in the bibliography to the WIID Documentation. The fourth series (row 2 column 3) comes from 'Jäntti 2005' (documented as unpublished estimates specially derived for WIID). According to *SurveySource2*, all four sets of estimates were derived from the Income Distribution Survey, except for the years prior to 1987 when the source was the Household Budget Survey. In addition, observe that there are six series labelled 'Per Inc Hou Hou' (see the Figure notes), with four of these missing information about *Curref*.

3.3 WIID: the need for explicit sample selection algorithms

Which observations for Finland might the analyst choose in order to have a consistent time series but only one observation per year? Clearly this depends on the purpose of the analysis (and hence choice of series) but, if a researcher wishes to study the distribution among individuals of household income equivalized by the modified-OECD scale, then the relevant series is the one in row 2 column 1. The other two 'Statistics Finland 2005' series refer to 'factor income' (row 1 column 3) and 'gross income' (row 3 column 2), with other dimensions the same. The 'Jäntti 2005' series is the same as the row 2 column 1 series, except that the equivalence scale is the per capita one rather than the modified-OECD scale. In fact, the single observation in row 1 column 4, for 1985, also uses the same definition as the row 2 column 1 estimates and also derives from the Household Budget Survey. (It comes from Atkinson et al. 1995, who cite estimates derived by Uusitalo.) Given this consistency, it may

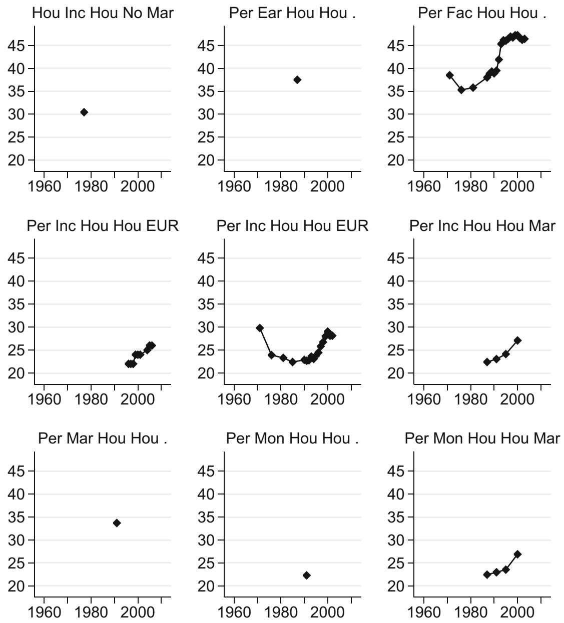


Fig. 2 Thirteen WIID series of 'Reported Gini' coefficients (%) for the distribution of income in Finland, 1960–2006. *Note* Only high quality observations are used, as defined in the main text. There is a separate graph for each series of estimates, with series defined using the fivefold classification discussed in the main text. The elements of each subtitle refer, reading left to right, to WIID variables *UofAnala*, *IncDefn*, *IncSharU*, *Equivsc*, and *Curref*, with '.' meaning that information is missing. Series with apparently identical titles differ in terms of either the publication that the estimate was drawn from or the original data source (WIID variables *Source1* and *SurveySource2*), or both

be appropriate to incorporate this observation into a single Finnish series, especially since there is no 1985 estimate from the 'Statistics Finland 2005' series.

Graphs like Fig. 2 can be produced for other countries. Atkinson and Brandolini (2001: Fig. 3) considered the case of the Netherlands though, by contrast with the Finnish case above, many of the 12 series shown were very short: 'two of the graphs consist of a single point; and five consist of only two or three points. The user is left with the ...problem of not knowing how to piece together the information in a meaningful way' (2001: 781). A similar set of graphs for the Netherlands and other countries derived from WIID (not shown) demonstrates that the problem of selection and splicing of series remains.

Thus, regardless of whether a WIID series is long or short, a user must inevitably do some forensic investigation of each series in order to select observations for analysis and, for consistency's sake, systematically employ some sort of selection 'algorithm'.

This is further illustrated by even seemingly straightforward exercises such as cross-national comparisons of inequality in a narrowly-defined range of years. In this exercise, I also benchmark WIID estimates against estimates from the Luxembourg Income Study Key Figures (LIS Cross-National Data Center (LIS) 2014) and from the Eurostat online database (Eurostat 2014). I use these particular benchmarks because both sources produce cross-nationally harmonized series from original data sources using a consistent set of definitions. In both databases, income is disposable income, with the sharing unit being the household and the unit of analysis is the individual. I selected WIID series that used these definitions

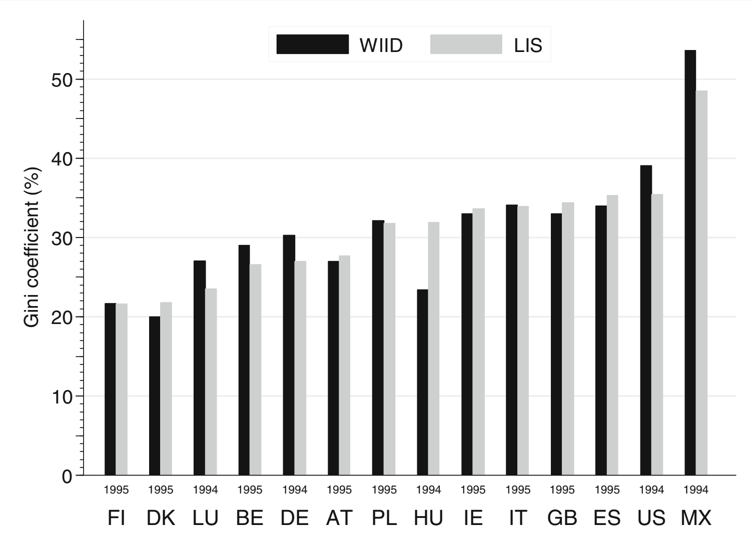


Fig. 3 Income inequality in the mid-1990s: WIID and LIS Key Figures estimates of Gini coefficients (%) compared. *Notes* LIS Key Figures estimates from LIS Data Center (2014), with the income distribution referring to household disposable (net) income among individuals, equivalized using the square-root-of-household-size equivalence scale. WIID observations refer to the Reported Gini and high quality observations only, and were selected using the algorithm described in the main text. WIID income definitions are discussed in the text. Countries are ordered left to right by the LIS estimates

too, in order to maximize comparability. The LIS Key Figures series employs a square-root-of-household-size equivalence scale; Eurostat employs the modified-OECD scale. In my initial selection of WIID observations, I did not restrict the equivalence scale definition, supposing that this would be a less crucial selection.

My first comparison is for the mid-1990s using WIID and LIS estimates of the Gini coefficient, and is motivated by Atkinson and Brandolini's (2001: Fig. 1) comparison of Gini coefficients among high-income OECD countries in the early 1990s using Deninger and Squire (1996) and LIS based estimates. Restricting the selection of WIID observations to high quality data for 1994 or 1995 led to multiple observations per year for each of 5 countries and so, in order to ensure that I had one observation per country-year, I had to examine the data sources for each series in detail and to make a selection (see my do-files for the precise choices). The results are shown in Fig. 3, with countries ordered from left to right in terms of the LIS-estimated Gini coefficient.

There are estimates available for 14 countries compared with 16 in Atkinson and Brandolini's for rich OECD nations in the 'early 1990s'. Only ten countries overlap in the two figures. (My estimates include observations for Austria, Poland, Hungary, and Mexico but not for Australia, Canada, France, the Netherlands, Norway, or Sweden.) Corresponding WIID and LIS estimates are within a couple of percentage points of each other for only 8 countries with some marked differences in the remaining 6 countries. The largest difference is for Hungary, with the WIID estimate some 9 percentage points lower than the LIS one. Hungary is placed second-lowest in the country inequality ranking according to WIID estimate but around half way up according to LIS. For the other five countries, the differences in estimates are around three to five percentage points. In each case, the differences reflect differences in definitions (e.g. equivalence scales). The Gini coefficient for most high-income countries rarely changes by more than about two percentage points per year so, by comparison with this benchmark, the differences between series for six of the 14 countries are relatively large. Put differently, if a researcher had undertaken comparisons using only data points based on exactly the same definitions (including e.g. the same equivalence scale), the number of countries with suitable data would be much smaller.

Overall, however, the differences between series are less than reported by Atkinson and Brandolini, suggesting some improvement over time in data quality. This is reassuring but of course the result is partly a consequence of my using a data selection algorithm that imposed a relatively high degree of comparability from the outset.

Figure 4 repeats the exercise for 2000 but now adds in estimates from Eurostat (2014). The observation selection algorithm was the same as described in the previous paragraph, except that a country's selection required a non-missing observation for each of the three series in either 2000 or 2001. All ten countries selected are EU member states, by construction, and the data are for 2000. The general impression provided by the chart is that there is generally a close correspondence between the Eurostat and LIS estimates (with the exception of Italy for which the difference is some four percentage points). However, the WIID estimates are out of line with the Eurostat ones for at least four countries. To be sure, some of these instances are where the LIS and Eurostat estimates also differ from each other (Belgium and Italy), but also note the case of Luxembourg where the LIS and Eurostat estimates are the same but the WIID one is four percentage points greater. The country inequality

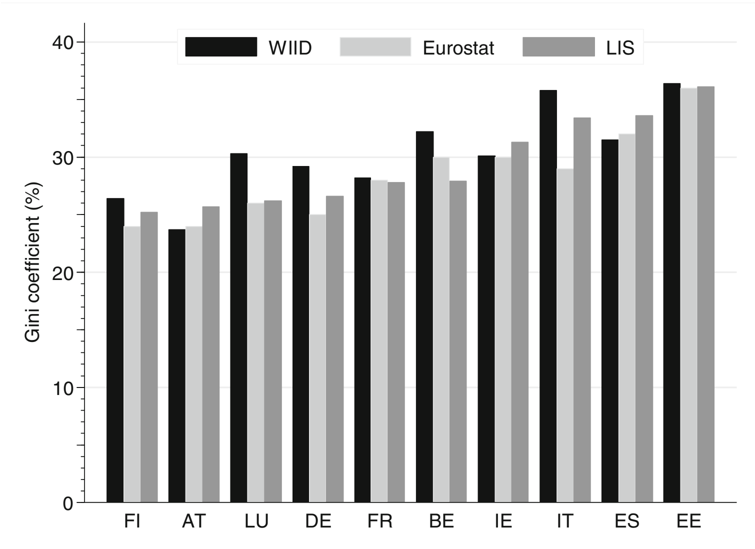


Fig. 4 Income inequality in 2000: WIID, LIS Key Figures, and Eurostat estimates of Gini coefficients (%) compared. *Notes* LIS Key Figures estimates from LIS Data Center (2014); Eurostat estimates from Eurostat (2014). In both series, income is disposable income, the sharing unit is the household, and the unit of analysis is the individual. The LIS Key Figures series employs a square-root-of-household-size equivalence scale; Eurostat employs the modified-OECD scale. WIID and LIS observations were selected using the algorithm described in the main text. The WIID estimates refer to the Reported Gini and high quality observations only. The income definitions underlying them are discussed in the text. Countries are ordered left to right by the LIS estimates

ranking according to WIID is rather different to that according to the other sources. Once again we see that one cannot simply use the WIID data 'as is' and the benchmarks provided by comparable data from other sources suggest a need to treat the WIID series with caution.

3.4 WIID: sample selection algorithms in action: the USA and China

The conclusion that differences in definition matter also applies to analysis of inequality trends over time, even for a single country. This is illustrated first by the case of the USA, which I choose because it has one of the longest single inequality series in WIID, there are other long US series with which the WIID estimates can be compared, and it is an 'important' country. The USA also illustrates the importance of issues of data processing. Look at Fig. 5, which displays four series of estimates of the Gini coefficient, all derived from annual Current Population Survey (CPS) data.

Users need to know, first, that the CPS has changed significantly over time. In particular there was a major redesign in 1992/3 that improved the collection of data on high incomes. Second, for confidentiality reasons, CPS data are top-coded and the censoring values have changed over time. The US Census Bureau has access to 'internal' CPS data in which the prevalence of top-coding is significantly less than the CPS data placed in the public domain ('public use' data); it is the internal data that underlie the Census Bureau's

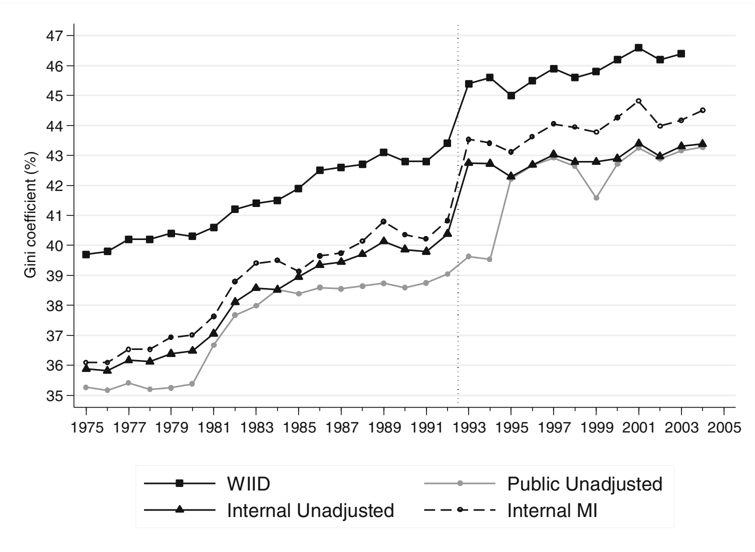


Fig. 5 Trends in the Gini coefficient for the USA: WIID and other series. *Notes* The WIID series refers to the Reported Gini for unequivalized gross household income with households as the unit of analysis. The other three series are taken from Burkhauser et al. (2011), and refer to gross household income equivalized by the square root of household size and with individuals as the unit of analysis. All four series are derived from Current Population Survey (CPS) data. The WIID series and the two 'Internal' series are based on CPS 'internal' data, in which the prevalence of top-coding is much lower than in 'public use' data. There was a major CPS redesign in 1992/3, and top code values changed throughout the period in both internal and public use data. The Public Unadjusted series includes US Census Bureau cell-mean imputations for top-coded observations from 1995 onwards. See the main text for more detailed discussion of the differences in definition between the series

published statistics. Moreover, from 1995 onwards, the Census Bureau replaced the top-coded values for high income observations in the public use data with 'cell mean' estimates derived from the internal data. So, although the measurement of high incomes in the public use data improved over time, there are major discontinuities related to 'processing' matters. Burkhauser et al. (2011) had 'special sworn status' access to CPS internal data and so were able to explore the consequences for estimation of the Gini coefficient of using different data series and different treatments of top-coding. The 'Internal Unadjusted' series refers to internal data in exactly the form used by the Census Bureau, except that the income definition is different (see below). The Burkhauser et al. (2011) 'Internal MI' series uses the same data except that the small number of top-coded observations in the internal data are replaced by multiply-imputed observations. This series provides the researchers' preferred CPS estimate of US income inequality trends over the period. The Burkhauser et al. (2011) 'Public Unadjusted' series uses instead the public use data as released by the Census Bureau, except for a change in the income definition described shortly.

The WIID series refers to the Reported Gini and is derived by splicing series from two US Bureau of the Census online sources, one dated '3 Feb 99' according to WIID (1975–1997) and the other '2/2005' (1998–2005). In both cases, the distribution is of gross monetary income with the household as the unit of analysis (not the individual), and there is no adjustment to money income using an equivalence scale. The definition is therefore quite different from the definition underlying the estimates shown for the USA in Fig. 3. The other three series also refer to gross monetary household income, but the unit of analysis is the individual (households are weighted by household size in the Gini calculations) and income is equivalized by the square root of household size. Thus, the WIID and Internal Unadjusted series use essentially the same data source, and differences in trends reflect differences in the definition of income. The Internal MI series also uses the same data source, but provides higher inequality estimates in any given year (as expected), with the magnitude of the upward adjustment depending on the prevalence of top-coding in the internal data.

The differences between the public and internal series, the changing treatments of top-coding, and the CPS redesign, are all data processing issues that complicate not only estimates of trends over time, but also inequality differentials between the USA and other countries. The congruence of the internal series and their difference from the public use series suggest the importance of using inequality estimates based on the Census Bureau internal data (as WIID does), but the cost is that the income definition is one that is not commonly used in cross-national comparative analysis nowadays. It might be argued that the remaining differences in income definition could be controlled for using dummy variable adjustments (see above). If such a procedure is to work in this case, the WIID and Internal Unadjusted series should move in parallel. But the absolute differences between the estimates range between 2.7 and 4.1 percentage points over the period; in proportionate terms, the WIID estimates range from being 6.2 % to 11.3 % larger than the Internal Unadjusted ones, with differences tending to be largest towards the beginning of the period. This variability raises questions about the appropriateness of simple dummy variable adjustments (or proportional adjustments as with the SWIID – see below).

Issues that arise with using WIID data for developing countries are illustrated by the case of China. China contributes a total of 121 observations, 120 of which refer to the period 1964 to 2004, and one to 1953 which I drop. Of the 120 post-1960 observations, 40 have *Quality* = 2 (the quality of either the income concept or the survey is problematic or unknown or the WIID producers have not been able to verify the estimates), and 80 have *Quality* = 3 (for observations where both the income concept and the survey are problematic or unknown). Many of the observations refer to urban or rural areas separately. There are only 34 for the country as a whole (*AreaCovr* = 'All'), and these are displayed in Fig. 6 separately according to quality rating.

There are 7 *Quality* = 2 observations spanning 1988 to 2003, and the remaining 24 *Quality* = 3 observations cover the period 1964 to 2004. The combinations of *UofAnala*, *IncDefn*, and *Equivsc* values characterize six different series (and all 34 observations refer to *AgeCovr* = All). This diversity is highlighted in Fig. 6, with a distinction made between observations for the unit of analysis is the person and the income definition disposable (the relatively small number with this consistent definition are marked with the filled squares and triangles). Observe the prevalence of multiple observations for some years: look at 1983, 1985, 1990, and especially 1995 where the difference between values is very large. Complicating researchers' choice of observation is the fact that the two observations for 1995 (both *Quality* = 2) are based on the same income distribution definition (disposable household income per capita, among persons), though derived from different surveys ('Sample Survey by the Economics Institute of the Chinese Academy of Social Sciences' and 'Rural/Urban Household Survey' according to *SurveySource2*) by two different research groups (see *Source1*). All in all, it can be a complicated business for researchers to select observations from WIID.

Researchers need to report their selection algorithms because different selection algorithms lead to different inequality series. This is illustrated by Xie and Zhou (2014: Fig. 1) who present a series of Gini coefficients for China for which WIID2c is cited as the source (though no selection algorithm is reported). Their series looks different from those shown in Fig. 6.

In sum, the WIID data for China illustrate a number of tricky issues for analysts of developing country inequality data. Data quality is poor relative to that for most rich countries, and restricting attention to observations from higher-ranked quality categories dramatically

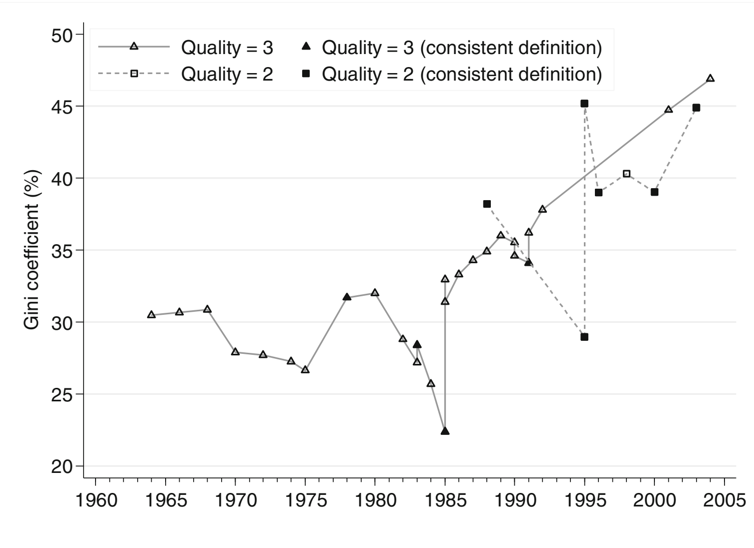


Fig. 6 Trends in the Gini coefficient for China, by WIID series. *Notes* The WIID series refers to the Reported Gini and is based on all observations with *AreaCovr* = 'All'. The subsets of observations with 'consistent definition' are those for which, in addition, *UofAnala* = 'Person' and *IncDefn* = 'Income, Disposable'

reduces the number of observations and time period covered. (To take a longer-term view in the Chinese case, a further compromise on data quality is required.) In addition, researchers have to face up to additional problems of non-comparabilities in income distribution. Given the various definitions and the quality of the data, it is difficult to assess in the Chinese case the extent to which the relatively large fall and rise in inequality during the early- to mid-1980s is genuine or a data artefact.

4 The Standardized World Income Inequality Database (SWIID4.0)

Frederic Solt's (2013*a*) summary description of SWIID on its homepage is as follows:

> Cross-national research on the causes and consequences of income inequality has been hindered by the limitations of the existing inequality datasets: greater coverage across countries and over time has been available from these sources only at the cost of significantly reduced comparability across observations. The goal of the Standardized World Income Inequality Database (SWIID) is to meet the needs of those engaged in broadly cross-national research by maximizing the comparability of income inequality data while maintaining the widest possible coverage across countries and over time. It standardizes the United Nations University's World Income Inequality Database, the OECD Income Distribution Database, the Socio-Economic Database for Latin America and the Caribbean generated by CEDLAS and the World Bank, Eurostat, the World Bank's PovcalNet, the UN Economic Commission for Latin America and the Caribbean, the World Top Incomes Database, national statistical offices around the world, and many other sources while minimizing reliance on problematic assumptions by using as much information as possible from proximate years within the same country. The data collected by the Luxembourg Income Study is employed as the standard. The SWIID currently incorporates comparable Gini indices of market and net income inequality for 173 countries for as many years as possible from 1960 to the present as well as estimates of uncertainty in these statistics. A full description of the SWIID and the procedure used to generate it is presented here ...
>
> Source: http://myweb.uiowa.edu/fsolt/swiid/swiid.html (Accessed 30 March 2014.)

The SWIID homepage also provides links to Solt's 2009 *Social Science Quarterly* article in which he describes the SWIID (an earlier version than the one under review) and how it is constructed, and to a Harvard Dataverse webpage from which the data and additional materials are available to download. Previous versions of SWIID are also downloadable (version 1 is dated September 2008). In terms of content and coverage, the main innovations in version 4.0 are threefold. There is the inclusion of additional observations so that the time period is extended through to 2012 for some countries (compared to 2006 in the WIID), and also top income share estimates (specifically the share of total income held by the richest 1%). Otherwise, the main difference from version 3.1 is the way in which the data are made available. In the earlier version, the dataset consists of Gini estimates and their associated imputation 'standard errors', and some separately-provided Stata do-file code illustrated how imputation uncertainty could be incorporated into estimation. In version 4.0, data in this format remain (in a Summary file), but the Main file is in a form that facilitates multiple imputation estimation techniques directly, as explained shortly. With all these features, it is clear that the SWIID is not a simple WIID adjunct, but can stand alone in a number of senses.

4.1 SWIID: data and documentation

Two zip files are downloadable. One contains the SWIID4.0 data, which come in two forms: there is a Main data file and a Summary data file. The Main data file comes in both Stata and R formats (each around 7MB). The Summary file is a file in 'csv' format (350KB) and so easily read directly by Excel, Stata, or R; it is intended to summarize the inequality estimates and their standard errors that are in the Main file (more about this later). The data zip file also contains a six-page pdf file on 'Using the SWIID'. The second zip file contains replication materials enabling users to reproduce Solt's work: there are data sets in spreadsheet form (WIID2c, plus additional income distribution summary statistics from the sources cited in the homepage statement above) plus a Stata do-file script and an R script that is called by it. I found the scripts essential for helping to understand various details of the database construction. Solt should be congratulated for his provision of replication materials; in this respect, all empirical researchers should emulate his example.

As with the WIID, I initially looked at the SWIID data in 'plug and play' mode without looking at any documentation and, again, I had to back off quickly and head for a more detailed consultation of the documentation materials.

The variable list in the Main file may appear puzzling to many users. (I refer to Stata versions of the data throughout.) There are six variables with names one might expect from reading the homepage statement such as the identifiers *country* and *year*, plus distributional summary statistics *gini_net*, *gini_market*, *redist*, and *share1*. (These are the Gini coefficients for net and market income, the percentage difference between them representing the proportionate reduction in inequality due to taxes and transfers, and the share of total income held by the richest 1%.) But there are also 100 additional variables for each of the four distributional statistics, each prefaced with $\_X\_$, where $X = 1, 2, 3, ...,$ 100. There are valid values for these variables (you can summarize them, for instance), but *gini_net* and the like have missing values for all observations. The key to understanding this is indicated by the presence of the variable *_mi_miss*. The dataset contains 100 multiply-imputed values for each distributional summary statistic, and *_mi_ miss* identifies this fact to Stata, so that users can directly apply multiple imputation versions of estimation commands.

An important conclusion from these initial explorations is not only that there are multiply-imputed observations in SWIID, but that *all* of the distributional summary statistics in the database are imputed. Put another way, no estimates from any data source, regardless of their quality, are left 'as is' in the SWIID (with the exception of LIS-based benchmarks used in a manner discussed below). I return to this issue later.

The 'Using the SWIID' document does explain that the data are multiply imputed, but the variable description on page 2 does not explain the variable naming conventions (it refers only to *gini_net* and the like, for instance). I suspect that users unfamiliar with multiple imputation methods (most economists?) may be put off or, at least, decide to work instead only with the Summary spreadsheet file. If you open that file, you see data in a much more familiar format: there is one country-year observation for each of *gini_net*, *gini_market*, *redist*, and *share1*. So, the Summary data are more directly useable, but employing them raises questions about how they relate to the multiply-imputed data in the Main file and what the consequences are of ignoring the multiple imputations.

Solt's (2013*b*) presentation, downloadable from the SWIID homepage, provides a user-friendly introduction to the SWIID and illustrates its use. It does not provide answers to many of the issues raised in this article, however.

4.2 An introduction to multiple imputation methods

In principle, the relationship between the data in the Summary and Main Files is straightforward according to the principles of multiple imputation (MI) analysis. MI consists of three steps. First, there is an *imputation* step that is repeated multiple times (which produces M datasets without any missing observations); second, there is *estimation* of statistics of interest using each of the datasets, and, third, *combination* of the separate analyses into a single set of MI estimates.

The combination of estimates at the third step almost invariably employs what is known as Rubin's Rules (Rubin 1987, 1996), which can be summarized as follows. Consider any scalar statistic, θ. Then, given M multiply-imputed data sets containing data to estimate θ, the point estimate of θ is the mean of the M estimates derived from each of the data sets. The variance of the estimate (standard error squared) is the sum of two terms. The first is the average of the M estimates of the sampling variance; this is the within-imputation variance summarizing sampling variability in each imputed data set. The second term is equal to $(1+ 1/M)B$, and summarizes imputation variability. The contribution of the between-imputation variance B is smaller, the larger that M is. The expressions for the MI point estimate and variance can be generalized to the vector case, and the estimators can be shown to have a number of desirable properties. MI methods are an improvement on single-imputation methods because they take into account the stochastic nature of the imputation process.

To make things concrete, suppose that we have a linear regression model for the Gini coefficient and θ is the coefficient on the explanatory variable of interest, and the model is fitted to a multi-country panel data set (as in Section 5). Multiply-imputed values of the Gini are available for each country-year (and assumed normally distributed), and no other variable is imputed. Researchers will get the same point estimates of the coefficient of interest, whether they run one linear regression using the average of the imputed Gini values, or use the appropriate MI estimation method based on the multiple regressions drawing on all of the imputed values. What will differ between the approaches is the estimated precision of the estimates. Using MI methods will lead to larger standard errors on coefficients – and hence less statistically significant estimates – because they take account of the imputation variability. (Some other differences between estimates from the two approaches may appear if researchers fit non-linear models, or the Gini coefficient is used as an explanatory variable rather than the dependent variable.) In Section 5, I investigate whether the effects on precision are large or small.

This short discussion of MI makes it clear that the principal role of SWIID is to enable researchers to skip the imputation step in their cross-country inequality analysis. Put differently, SWIID's analytical validity rests on the credibility of the imputation model that is employed since, given the multiple imputations, the second and third steps in the process are straightforward if you have access to suitable software. For example, in Stata, the `mi estimate:` prefix command implements the second and third steps for many types of estimation routine. Of course, for users who ignore the multiple imputations and simply employ the mean value for each country-year observation, the validity of the imputation step is just as important.

In the SWIID case, $M = 100$, and so one would expect the *gini_net* entry in the Summary file for a given country-year cell to be the average of the 100 imputed *gini_net* entries for the same country-year cell in the Main file, and similarly for the other distributional variables. As it happens, this is not the case for two reasons, as I shall explain below when discussing the contents of the database in more detail.

An important first step, however, is to explain the nature of the ‘imputation model’ underpinning the SWIID’s construction and to consider its credibility. Because there are no external benchmarks for the missing data (except in the rare case where new data have become available after the SWIID’s compilation), assessment of imputation model has to rely to a large extent on a consideration of the assumptions built into it. I set out the three stages in the imputation process, and then return to assess them.

4.3 SWIID’s imputation model: explanation

The first stage in the SWIID imputation procedure concerns *database inclusions*. The main dataset is WIID2c, as discussed above, but additional Gini estimates are incorporated from a Statistics New Zealand source and, much more importantly, from the LIS Key Figures (as of 13 June 2013 according to the SWIID do-file), together with a set of market and net income Gini coefficients derived by Solt from LIS unit record data, and top income share data from the World Top Incomes Database (WTID, Alvaredo et al. 2014; version of 13 August 2013).

The second stage is *exclusion of observations*. Observations that are not based on coverage of the whole population or all age groups are dropped (with the exception of some urban inequality estimates for Argentina and Uruguay), and observations referring to years before 1960 are also excluded on the grounds that they are ‘often based on unreliable surveys’ (Solt 2009: 235).

The third stage is the *imputation procedure* itself. As this is complicated, I will first provide an intuitive explanation and then discuss the intricacies. Suppose there are two data series for the Gini coefficient available for a large number of country-year observations, one based on gross income and the other on net income, but estimates are missing for the net income Gini in some cases. If one could assume that the ratio of the net income Gini to the gross income Gini were constant within some group g of country-year observations, and one had an estimate of that ratio, call it R_g, then one could impute the missing values. The net income Gini imputation for a missing country-year observation within group g is equal to its observed gross income Gini multiplied by R_g.

SWIID uses regression methods to estimate R_g for each of a number of groups of country-year observations. Net-to-gross Gini ratios are calculated for all country-year observations with non-missing values and these are then regressed on country-year-group variables such as country-decade or region. (Each set of country-year cells characterized by these covariates constitutes a group assumed to have a constant ratio.) R_g is derived as the prediction from the fitted regression equation, and the uncertainty associated with the imputation is summarized by standard error of the prediction, assumed to be normally distributed with mean zero. Gini imputations are derived by multiplying together the derived ratios and observed Ginis. Estimates for the uncertainty of the Gini prediction can also be derived. Repeated draws from the normal distributions of predicted Ginis yield multiple imputations.

4.4 SWIID’s imputation model: complicated but important details

In practice, the derivation is much more complicated than as described. There are around twenty ‘data types’ (data series) distinguished, not two, meaning that many more different combinations of cross-series ratios are available. SWIID’s imputation procedure works with all possible combinations. Most of the data types are characterized in terms of definitional differences (using WIID or similar variables), but there are also separate series reserved for estimates derived from LIS Key Figures and top income share data from the WTID. (Although only imputations for the share of richest 1% are output, data on the shares of the richest 10% and richest 10% are also included in the imputation routine.) To ensure the top income shares are in the same metric as the Gini coefficients, shares are transformed into 'Pareto-Ginis' using standard formulae for a Pareto distribution and then transformed back to the share metric at the end.

The regressions for ratios use different sets of explanatory variables to characterize observation groups depending on how many non-missing country-year observations on ratios are available in each of the relevant series combinations. In the most data-rich situation, the regressions use country-decade indicator variables: ratios are assumed constant within each country-decade combination. With fewer data, ratios are assumed constant with each country, or within each of eight world regions (country groupings defined by Solt) or, finally, in the most basic case, within 'advanced' and 'non-advanced' nations. Unfortunately one cannot tell from the do-file code which particular grouping definition is applied to different country-year observations when the do-file code is run.

The predictions of missing ratios and associated prediction errors derived from these regressions are what SWIID calls 'one step' imputations. In addition, 'two step' predictions are made, exploiting the fact that a ratio for a pair of series *a* and *b* can also be written in terms of the ratio for *a* and the LIS series and the ratio for *b* and the same LIS series. The motivation is that 'for some combinations of *a* and *b*, there are few or no observations of the Gini index available, making ... one step [prediction] impractical or impossible' (Solt 2009: 236). Solt states that it may also lead to lower prediction error.

Moreover, in parallel, a set of non-linear loess regressions is run on the time series of ratios for each country separately (if there are more than three observations per country). The fitted parameters from these regressions are used to predict (interpolate) missing ratios and their uncertainty is summarized by the standard error of the prediction, as above. SWIID replaces the one-step regression imputations with their loess counterparts if the latter are available and also have a smaller prediction error. The resulting estimates are then replaced by corresponding two-step estimates if they are available and have a smaller prediction error.

The next step is to generate predictions of Gini coefficients (for net and market income and also top income shares) from the predictions of ratios. These estimates and associated prediction errors are then compared with corresponding estimates and standard errors derived from LIS unit record data on net income and, again, the estimates with the smallest uncertainty are the ones retained. It is essentially this comparison with 'gold standard' LIS data that leads to the 'Standardized' label in the SWIID's name. My view is that SWIID would be better labelled '*mi*WIID': it is the *m*ultiple *i*mputation nature of the data that are its truly distinguishing feature, and the pun is intended because a specific imputation model (Solt's) underpins SWIID.

The penultimate step in the imputation procedure is to generate 1000 simulated values for each country-year observation in each of the retained series assuming that each series is normally distributed with standard deviation given by the prediction standard error.

The final step is to smooth and interpolate each of the simulated series using a five-year moving average algorithm, implemented on the grounds that 'the distribution of income within a country typically changes slowly over time' (Solt 2009: 237). Two types of observation are excepted from this: the first are those derived from LIS unit record data, on the grounds of their high quality; the second are those referring to countries in Eastern Europe and the Soviet Union over the period when communist rule collapsed (on the grounds that large changes are likely in this scenario). Finally, the *redist* series of multiple imputations are derived from the imputations for *gini_net* and *gini_market* so that imputation variability is correctly accounted for in its calculation.

The SWIID imputation procedure generates 1000 imputed values per series with only the first 100 values of each series released in the Main File. But the mean values of each series that are placed in the Summary File are derived from the full 1000 values. This is confusing at the very least and I only discovered it by close inspection of the Stata do-file code. To be sure, one would expect corresponding means from the Main and Summary files to be similar given the assumption that imputed values are normally distributed. I show later that this is generally the case – except for the estimates of the income share of the top 1%, for which there are some very large differences.

4.5 SWIID's imputation model: commentary

It is clear from the preceding discussion that SWIID's imputation procedure is remarkably complicated and the details will be opaque to most users even after a close reading of Solt (2009) or his do-file. On these grounds alone, analysts may wish to avoid SWIID. (Should one use data if one doesn't know how they are derived?) Improvements to and elaboration of the documentation and replication materials may help mitigate this issue. Which particular top income series from the WTID is used for each country is not documented, for instance.

More importantly given the 'standardization' goal in SWIID, users are not provided with details of how the Gini coefficients from the gold standard LIS unit record data are derived, in particular the definitions of income distribution employed, the treatment of low and high income outliers, how standard errors were estimated, etc. Also unclear is how the imputation procedure distinguishes between, and differently uses, data from the LIS Key Figures and derivations from the LIS unit record data. (My detective work suggests that the LIS Key Figures – with distribution definitions as described earlier in this article – are employed for the point estimates of net income Gini coefficients, but LIS unit record data were used in a separate exercise to derive standard errors for these, as well as for the market income Ginis and their standard errors.) It is not stated explicitly that the 'standardization' means that the net and market Ginis are estimates for distributions of equivalized household income among individuals, where incomes are equivalized by the square root of household size (the LIS Key Figures definition).

In addition, users have no way to ascertain the importance of different components of the imputation procedure. For example, what proportion of observations is generated by the within-country loess regressions rather than the country-year regressions, and what proportion derive from one-step or two-step predictions of ratios? What are the precise definitions of the groups within which Gini ratios are assumed constant, and how many country-year observations are there within each group? How many observations are imputed at the final interpolation and smoothing step? A relatively straightforward way to help address issues such as these would be to add code to the Stata do-file so that it provides the relevant summary statistics, and to also provide users with the log-file that is produced by the do-file. (No log-file is included in the replication materials.) The documentation could then refer to the information in these materials.

I have quibbles with the characterizations of the data types (series) employed in SWIID. Choices have to be made because there is a large number of possible combinations of definition (see the WIID discussion above), but Solt's (2009: 235) discussion of the 'reference unit' confusingly combines the separate dimensions of income-receiving unit, unit of analysis, and equivalence scale.

SWIID's imputation procedure makes no use of the WIID's data quality assessment variable. Restricting attention to observations from 1960 onwards removes many low quality observations, but not all of them, especially for developing countries. I would have imagined that there was some way of taking advantage of the quality assessments in the imputation process directly. For example, it is rare for data producers responsible for surveys to impute all observations. More commonly, item non-response imputation – an activity quite similar to the SWIID's – retains each non-missing observation 'as is' and imputes only the missing values. (One difference with the SWIID is that it not only imputes but also standardizes relative to LIS Key Figures benchmarks.)

More generally, it would help users if the SWIID documentation related its imputation procedures to those typically employed by data producers and researchers. This discussion would cover not only SWIID's stochastic linear regression prediction method (contrasting its properties with those of procedures based on matching such as the hot deck, for example), but also the way in which the SWIID's combination of one-step, two-step, and loess regressions compares with methods commonly used to impute multiple missing variables sequentially such as 'chained equations' (Raghunathan et al. 2001).

I am not convinced by the application of the moving average smooth and imputation at the final stage of the imputation procedure. Although income inequality changes between one year and the next are not large in most countries, this is a yardstick by which to judge whether an imputation model is producing reasonable estimates rather than a property that should be imposed *ab initio*. Moreover, many inequality changes may well be true step-changes reflecting, for example, changes in the income tax rate structure or benefits policies, short-term macroeconomic fluctuations, or a major survey redesign, in which case smoothing out the effects of such changes is inappropriate. A concrete illustration is given below.

A more fundamental question concerns the assumption employed at the heart of the imputation procedure, specifically the assumption of constancy of ratios of Gini coefficients across data series within groups of country-year observations. It is a multiplicative version of the dummy variable adjustment procedure discussed earlier that assumes constant absolute differences between series. SWIID's assumptions are not as strong, because ratios are allowed to differ between groups of observations rather than being the same for all country-year observations, though observe that the dummy variable adjustment method can be straightforwardly extended to allow for variation in Gini differences across countries and time by using interaction variables in addition to intercept shifts.

Solt (2009: 233) refers to dummy variable and related adjustment methods and concludes that they are problematic, also stating later in his article that:

> ... as noted previously, the relationship between Gini indices with different reference units and income definitions will vary considerably from country to country and also over time depending on the extent of redistributive policies, details of tax law, patterns of consumption and savings, family structure, and other factors. In other words, ρ_{ab} is not constant but varies across countries i and years t. (Solt 2009: 236.)

Nonetheless, the gains from Solt's approach relative to simple dummy variable adjustments are hard to assess because the variation in ratios across countries and time (ρ_{ab} in his notation) that is actually implemented in SWIID is not apparent (see above).

Clearly, assumptions of constant differences and constant ratios (albeit within groups) are convenient, but I am not convinced that the SWIID implementation is sufficiently credible to bear the weight that is placed upon it.

Earlier I cited evidence about patterns of differences between market and net income inequality changing over time and across countries (OECD 2011: Chapter 7). Assuming constancy of ratios even within country-decades is implausible to me. See, for instance, the US evidence for changes in the gap between income series with different assumptions about the recipient unit and equivalence scale that I presented earlier. The assumption is even less plausible for regional groups. I would not expect to see the same constant-ratio relationship in, for example, China, Indonesia, Bangladesh, Turkey, and Fiji (all part of a 24-member Asian region), or among the 33 countries in the Caribbean, Central and South American group.

The essential problem is that there are two competing demands that cannot both be met. On the one hand, country-year observations have to be grouped in order to have donor observations to provide the values to be imputed to the missing observations and, other things being equal, the larger the group size, the more reliable is the within-group mean used for the imputation. On the other hand, there should be as many groups as possible to allow for the acknowledged variation in Gini ratios but, other things being equal, having more groups means a smaller average group size and, in the limit, no potential donor observations. The inevitable but unfortunate situation given the available source data is that groups are relatively broadly defined in SWIID, and so the assumption of within-group constancy in Gini ratios is very likely to be compromised. The same is, of course, likely to be true for Gini differences, which means that regression-based adjustments to WIID data for differences in variable definitions need to more sophisticated than simple intercept shifts.

It would be useful to gather more extensive evidence about Gini ratios and differences, covering the full range of differences in distribution definition and for a large number of countries and time periods. This would inform the use of both dummy variable adjustments with WIID data and the ratio adjustments that underpin the SWIID's imputation model.

4.6 SWIID: coverage and content

I turn now to discuss the contents of SWIID in more detail. Coverage is summarized in Table 3 with a breakdown of numbers of country-year observations by region and period. The total number of observations, 4,597, is around 87 % of the total number in WIID (Table 1, top panel). None of the observations refers to years before 1960 by construction,

Table 3 SWIID: number of country-year observations, by geographical region and year

Region	Period						
	1960–1969	1970–1979	1980–1989	1990–1999	2000–2006	2007–2012	*Total*
Africa	36	61	157	328	274	78	*934*
Western Europe (EU15)	71	101	151	160	112	82	*677*
Other Europe, Turkey, Russia	40	65	174	249	192	127	*847*
North America	20	20	20	20	14	10	*104*
Central & South America	33	81	180	259	165	87	*805*
Central, East, & South East Asia	61	100	214	247	184	105	*911*
Oceania	20	23	30	35	33	14	*155*
Middle East	4	14	35	51	44	16	*164*
Total	*285*	*465*	*961*	*1,349*	*1,018*	*519*	*4,597*

The numbers of observations refer to country-year cells. For each cell, there are 100 multiply-imputed values in the SWIID Main file and a single value (the mean of the imputations) in the Summary file and around 11 % refer to the period 2007–2012 (not covered by WIID). The consequence of the imputation procedure is that, by comparison with the subset of high quality observations in WIID (Table 2), there is substantial coverage of regions such as Africa and Asia, each contributing 20 % of the total number of observations. Only 20 % of SWIID observations are from EU-15 member states. Thus, if one is willing to trust the SWIID imputation model, the database offers substantial scope for global analysis.

In what follows I concentrate on 'advanced' countries, however, mainly in order to provide a closer contrast with my analysis of WIID data. Discussion of non-advanced economies is relatively brief. Also, I focus on the multiple imputation aspects of SWIID as they are its distinguishing feature. To this end, I first display the distributions of multiply-imputed net income Gini coefficients year by year for a small number of countries. All these estimates are derived from the SWIID Main file.

The MI estimate of each country-year Gini coefficient is equal to the mean of the 100 imputed Ginis (see the discussion of Rubin's Rules earlier). However, a conventional MI estimate of its standard error cannot be derived. Each country-year observation is a singleton, and the within-imputation variance referring to sampling variability (an average over M–1 observations) is undefined when $M = 1$. (Stata's `mi estimate: mean gini_net if country == ''C'' and year == ''Y''` for country C and year Y returns a missing value for the SE of the mean.) Solt (2013b) summarizes between-imputation uncertainty in terms of what he calls the '95% confidence interval', which he calculates using the standard deviation of the imputed estimates for each country-year observation and an assumption of normality. My view is that reference to a confidence interval in this context is potentially misleading (it differs from the conventional use of the term, which refers to sampling variability). I prefer to summarize the distribution of imputed country-year values of the Gini coefficient by plotting all of the values as well as the mean. MI standard errors for estimates exist when country-year observations are pooled, such as in a regression analysis based on multiple years or multiple countries (or both): see below.

Look first at the case of Finland displayed in Fig. 7, and recall the many WIID series for the Gini coefficient in Finland shown in Fig. 2. The black line connects the yearly means of the MI-estimated Ginis (the values used by researchers who ignore the multiply-imputed nature of the data), and each imputed value is shown as a gray dot. Greater imputation uncertainty means that there is a wider range of estimates around the mean for a given year. For reference, the seven Gini estimates from the LIS Key Figures are also shown (I have used the data available in February 2014; Solt's LIS data are from June 2013).

It turns out that the U-shape pattern, including the noticeable flattening out around 2000, traced out by the mean SWIID estimate corresponds reasonably closely to the patterns shown by the 'preferred' WIID series. (The series from Fig. 2 row 2 column 1, based on a similar income definition to the SWIID's net income one, is reproduced in the figure.) Differences between the series in the mid-1970s are difficult to assess because there are many fewer WIID observations in this period but, generally speaking, the levels of estimated inequality are broadly similar. The difference in Ginis in corresponding years is at most around two percentage points, though note that the WIID observations often lie outside the range of the SWIID observations. Moreover, the WIID observations (which are of high-quality) suggest a greater increase in inequality in the late-1990s than do the SWIID ones.

The imputation variability in the estimates is relatively small between the mid-1980s and the early-2000s, which is precisely the period in which the frequency of high quality WIID observations is greatest. There were no WIID observations prior to 1965 in the high quality observation sample used for Fig. 2 (not shown in Fig. 7) but observations for this period

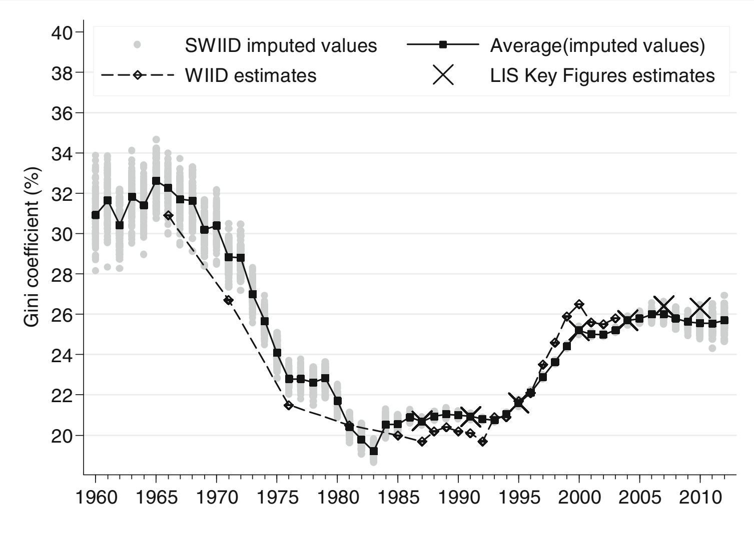


Fig. 7 SWIID estimates of the net income Gini coefficient for Finland. *Note* The WIID estimates are those shown in column 1, row 2 of Fig. 2 (see main text for further explanation)

have been imputed by SWIID, albeit with a relative large degree of imputation variability. There are three WIID observations on the Reported Gini for 1962, each around 47 % which is well beyond the upper range of the SWIID imputations for that year (around 32 %). On the other hand, each corresponding WIID observation is of low or indeterminate quality (*Quality* = 3 or 4) and the income definitions are quite different from those implicit in the SWIID estimates.

The SWIID series are 'standardized' with reference to LIS estimates, and Fig. 7 shows that the mean of SWIID estimates coincides with the LIS Key Figures net income estimate of the Gini coefficient for five of the seven comparisons. For the remaining two years (2007, 2010), inequality is underestimated by the SWIID relative to the benchmark series, though the difference is small in percentage point terms and within the range spanned by the SWIID imputations (though that nuance is lost if researchers use only the mean value). Nonetheless, it is interesting that there is any difference at all; none is apparent for the other countries shown in the next three Figures. The reason for the difference is that LIS Key Figures estimates for 2007 and 2010 were not available when SWIID was compiled in 2013, so this a rare occasion in which the SWIID's out-of-sample prediction can be checked.

The SWIID estimates for the USA are shown in Fig. 8, together with two of the US series shown in Fig. 5. When comparing the series, remember that the earlier estimates refer to distributions of unequivalized household gross income among households rather equivalized household net income among individuals. SWIID imputation variability is greatest in the pre-1975 period, which is when the WIID series changed its source (and hence is not shown for that era). It is primarily the difference between gross and net income that explains the difference in Gini levels. By construction, the LIS Key Figures estimates and the means of the SWIID estimates coincide in all ten years in which comparisons are possible.

However, comparisons of the various series suggest some problems with the SWIID imputations. For example, the SWIID estimates suggest that income inequality fell between 1975 and 1980 and then rose to its 1975 level again by 1984. This contrasts with the pattern

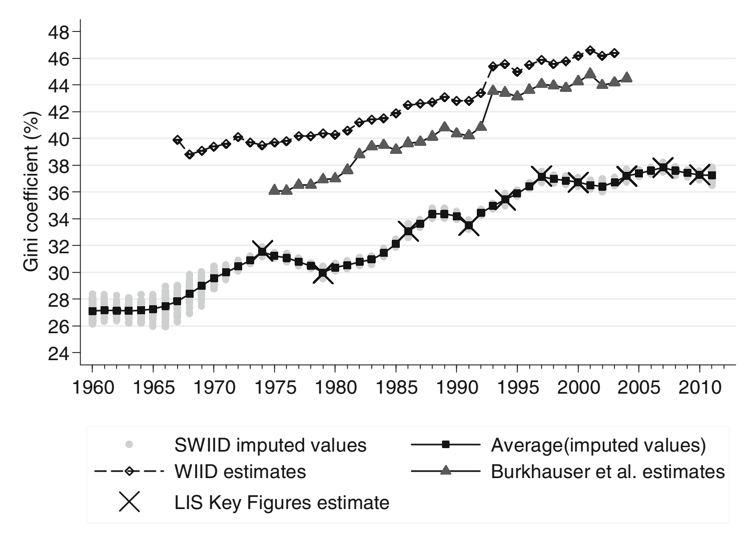


Fig. 8 SWIID estimates of the net income Gini coefficient for the USA. *Note* The WIID and Burkhauser et al. estimates are those shown in Fig. 5 and are based on a gross income definition whereas the SWIID estimates are based on a net income definition (see main text and notes to Fig. 5)

shown by the four series shown in Fig. 5 (two of which are reproduced in Fig. 8). For example, the WIID and Burkhauser et al. (2011) estimates, all based on Census Bureau 'internal' data, indicate a gradual rise over the first five years followed by a sharper increase in the subsequent four. There is also the issue of how the effect of the major CPS redesign in 1992/3 is handled. Figure 5 shows the sharp discontinuity correctly; the SWIID series in Fig. 8 does not, most likely because of the moving average smoothing algorithm employed in the imputation procedure.

Britain is a country for which there are long and consistent series of annual observations on the Gini coefficient available in the WIID and national sources, and so provides a good opportunity to examine how SWIID estimates compare with other reference points. See Fig. 9. The SWIID estimate for each year is the mean of the imputed values for that year (the full range of imputation values is suppressed, for legibility). The WIID series is derived from estimates provided by Institute for Fiscal Studies (IFS) spreadsheets dated March 2004 and April 2006. I took the series labelled IFS from the most recent spreadsheet at the time of writing (Institute for Fiscal Studies (IFS) 2013). (All three spreadsheets were produced to accompany annual IFS reports on inequality and poverty.) The income distribution definition in the WIID and IFS series refers to equivalized household net income among individuals. The difference between the series is that the equivalence scale in the former case is the so-called McClements scale and it is the modified-OECD scale in the latter case. As Fig. 9 shows, the effect of this difference is minor: the Gini coefficient is less than a percentage point greater than the IFS series in corresponding years, and the two series move in parallel. These series provide important benchmarks because the data are widely acknowledged to be of high quality, and they use definitions that are exactly the same as those used since the mid-1990s by the UK Department of Work and Pensions' official statistics on

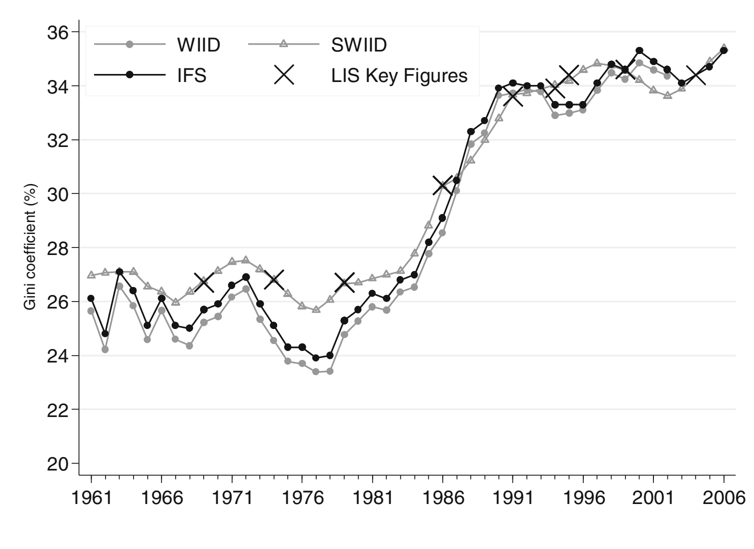


Fig. 9 SWIID estimates of the net income Gini coefficient for Britain. *Notes* The SWIID estimate for each year is the average of the 100 imputations for that year. (The full distribution of SWIID imputations for each year is not shown, for legibility.) The WIID and IFS series refer to estimates derived by the Institute for Fiscal Studies. See the main text for more discussion of the sources and income definitions

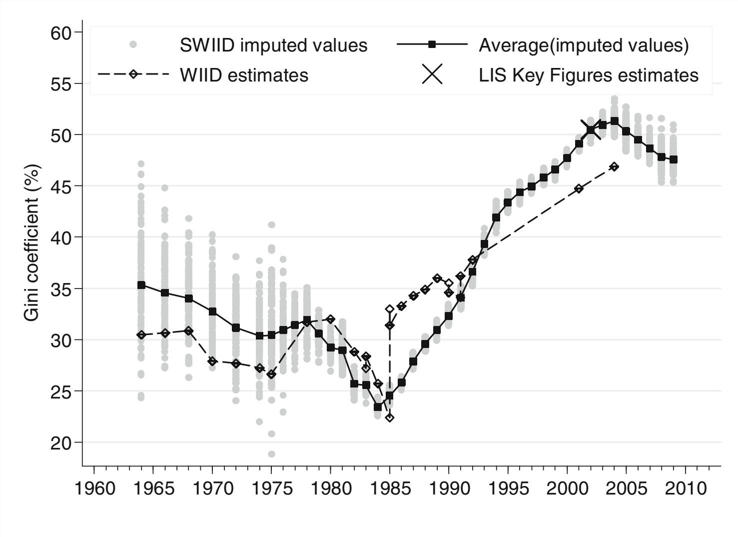
income distribution, and the survey data sources are the same as well. The data and definitions have been subject to much scrutiny by the department, the IFS, and other researchers, and the estimates and year-on-year changes in them receive much media attention.

Given this background, the SWIID estimates seem problematic in several respects. Although the LIS series on which it is standardized uses a very similar income definition to those for the other series, it is surprising that the gap between them is not constant over time. The SWIID series lies above the IFS series until the mid- to late-1980s (with the difference varying but often more than one percentage point) but, thereafter, the SWIID series weaves in and out of the IFS and WIID series. Also, the SWIID series is too smooth by comparison with the high quality benchmarks. For example, it under-estimates the rate of increase in equality between 1977 and 1990, and the fall in inequality in the early 1990s is not picked up. Some might argue that these differences are relatively small, but differences of one to two percentage points in Gini coefficients are non-trivial by comparison with what are considered to be the limits of changes between one year and the next. (I am ignoring issues of whether the estimates differ from the point of view of statistical significance; here the issues concern data per se rather than sampling variability.)

Readers may be impressed by the relatively small amount of imputation variability illustrated by the cases of Finland and the USA in Figs. 8 and 9. After all, the range of Gini imputations for a given year is at most around 6 percentage points (Finland in the early 1960s) and often much less. However, this impression is potentially misleading.

Imputation variability for some countries included in SWIID, especially developing nations, is huge. This is illustrated by the cases of China and Kenya shown in Fig. 10. The range of SWIID imputed values for the net income Gini for China is around twenty percentage points for the decade prior to 1975. For Kenya between 1960 and 2005, the range is never less than 10 percentage points, is often at least 20 percentage points, and reaches a maximum of 75 percentage points (in 1964). These cases raise the question of what the

(a) China



(b) Kenya

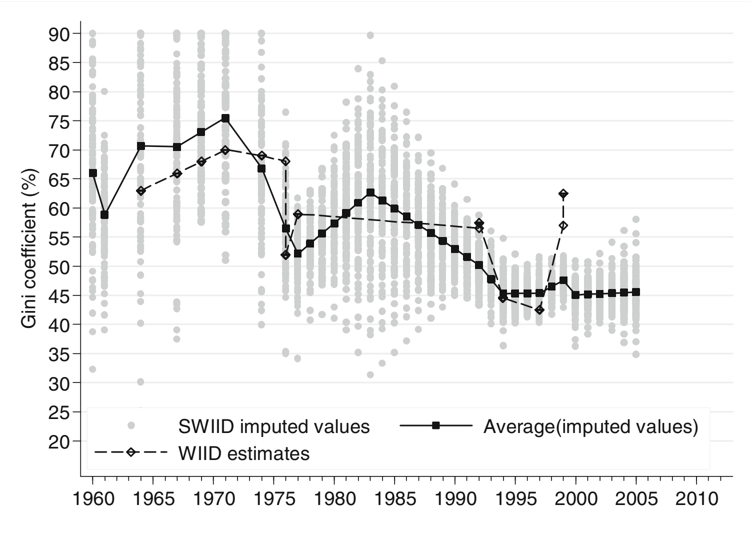


Fig. 10 SWIID estimates of the net income Gini coefficient for China and Kenya. **a** China **b** Kenya. *Notes* The WIID estimates shown for each country are based on all observations with *Quality* = 3 and *AreaCvr* = 'All'. All other WIID observations for Kenya are of poorer quality. The shorter *Quality* = 2 WIID series for China is shown in Fig. 6

impact is of imputation variability on the precision of estimates. I address this question in the next section.

The SWIID tendency to smooth out changes over time that was remarked on earlier is apparent in Fig. 10 as well. The figure also shows how the SWIID adds new observations where WIID data are missing and also beyond the period spanned by the WIID observations.

Moreover, the SWIID imputations for both China and Kenya for the very end of the period indicate an inequality trend that is quite different from the trend that would be derived by naïve extrapolation of the WIID series. This raises questions about either the validity of the SWIID imputation model in these cases or the quality of the WIID observations (or both). Further support for the first position comes from Xie and Zhou's (2014) detailed study of income inequality trends in China. Focusing on the period since the mid-2000s, they use estimates from multiple household surveys to make a persuasive case that inequality continued to rise after 2005 (it is the poorer quality 'official' estimates that show a decline in inequality over this period).

4.7 SWIID: estimates of the share of the top 1%

I turn now to the problems with the SWIID estimates of the share of the top 1% that I alluded to earlier, namely that the country-year means of imputed values in the Main File differ significantly from the country-year means that are provided in the Summary File. The problem is illustrated by Fig. 11, which uses data on all 4,597 country-year observations on the share of the top 1% in SWIID. The vertical axis shows the share derived from the Summary File; the horizontal axis shows the share derived by taking the mean of the 100 imputed values for each country-year observation in the Main File. The points should all lie very close to or on the 45°ray from the origin (with the only differences arising from taking means over 100 imputations rather than 1000 imputations: see earlier). The problem is that there is a significant fraction of country-year observations for which the estimate from the Main File is substantially greater than the corresponding estimate from the Summary File: the points on the lower right hand side of the figure represent 857 country-year observations from 25 countries (including for example Finland, the UK, and the USA).

I have not been able to fully understand what generates the inconsistencies. The problematic observations appear to relate to periods during which there is a break in the relevant WTID data series that SWIID draws on. The fact that the problematic observations trace out

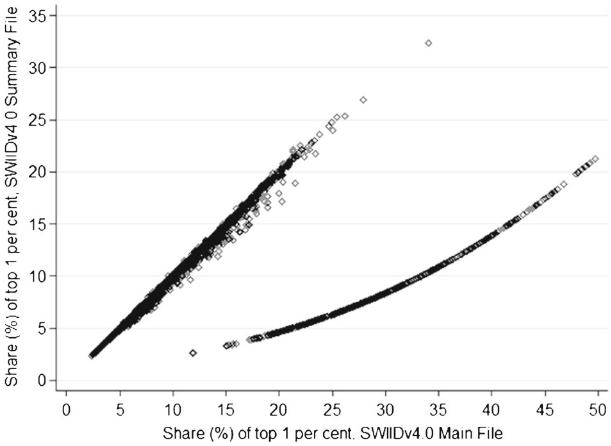


Fig. 11 SWIID estimates of the income share of the top 1%: scatter plot of Summary file estimates against Main file estimates. *Notes* Each point in the figure shows estimates of the share of the top 1% for each of the 4,597 country-year observations in SWIID. The vertical axis shows the share derived from the Summary File; the horizontal axis shows the share derived by taking the mean of the 100 imputed values for each country-year observation in the Main File. The points in the lower right hand side of the figure represent 857 country-year observations from 25 countries

a smooth curve suggests that the Pareto-Gini transformation that SWIID uses (see above) is also playing a role. But I have not yet discovered why the apparent problem does not show up in the Summary File as well as in the Main File, a puzzle since the former is derived from the latter. (The Summary File entries appear to be fine.) Note, by way of contrast, that there are no inconsistencies between Summary and Main File estimates for net or market income Gini coefficients. In each case, their distributions are almost identical, with Pearson correlations greater than 0.999. On the basis of this analysis, I recommend that the top 1% share imputations be removed from SWIID until the problem is resolved. At the very least, a warning should be posted on the SWIID website. There are also more fundamental questions about whether top income shares can be imputed using the same procedures as Gini coefficients for the distribution as a whole.

The estimates of the share of the top 1% are also an important reminder of the large extent to which SWIID may fill in missing observations. Iceland and Hungary are countries that do not appear in the WTID, but SWIID provides imputations for Hungary for 1967 and then annually from 1981 through 2001 and for Iceland annually for 1992 through 2011. Data for the share of the top 1% covering 2009–2011 were added to the WTID after SWIID4.0 was released. The SWIID does contain estimates for these years, however, and the 95% CI for the share of the top 1% for each of the last two years does not contain the corresponding WTID estimates. This points to potential problems with the SWIID imputation model for top income shares.

5 Illustrative regression analysis: WIID and SWIID

Atkinson and Brandolini (2001, 2009) showed that the use of different inequality series, samples, and explanatory variables – including dummy variable adjustments – could have a significant impact on the findings derived from econometric work. I re-examine this issue using inequality data from WIID and from high quality sources (LIS and Eurostat). I also illustrate the consequences of making different choices using SWIID data (using samples for which there are not full sets of benchmark observations). Observe that the intrinsic validity of the regression point estimates cannot be assessed for samples with global coverage: by definition there are no external benchmarks provided by models based on unobserved missing data. However, one can examine the impact on estimation precision of the uncertainty introduced by the multiple imputation procedure, and hence whether analysts are likely to draw the wrong conclusions about statistical significance if they ignore the multiply-imputed nature of SWIID data.

I address these various issues in two ways. First I analyze the relationship between income inequality and the macro-economic factors such as unemployment and inflation, looking at WIID and SWIID data in turn, but focusing on advanced economies. Second, with SWIID data, I use simple country-specific regression models of inequality trends to examine the impact of imputation variability, and I include both rich and poor countries in the analysis.

5.1 WIID-based regressions of the relationship between income inequality, unemployment, and inflation

The inequality-macro literature is often commonly associated with the pioneering research of Blinder and Esaki (1978). Although their regressions used quintile group income shares as dependent variables, there is also a substantial literature that uses the Gini coefficient, and I follow that practice here. According to Parker (1998, especially Table 2), most studies of this type have found that a higher Gini coefficient is associated with lower inflation and with higher unemployment, though he also comments that all but two of the twelve studies reviewed used time series data from the US CPS. (Parker also discusses various econometric issues raised by such analysis – ignored here. See also Jäntti and Jenkins 2010 on this topic.)

My data on inflation and unemployment rates are from the IMF's World Economic Outlook 2013 database (International Monetary Fund (IMF) 2013), restricting attention to years from 1980 onwards and to 'advanced countries' in order to impose some homogeneity. For the purposes of this assessment, I have taken the WEO data 'as is', even though they are another secondary data collection and deserve scrutiny in the same way that WIID and SWIID data do.

My WIID selection algorithm was to first choose country-year observations for countries and periods in which the income definition referred to gross or disposable income. This yielded 727 country-year observations for 21 countries (listed in the notes to Table 4). There were 12 countries with multiple country-year observations, necessitating detailed inspection of each country's data series in order to select observations. Wherever possible, I chose for each country the ones providing the longest series according to a particular definition. After this selection, there were 242 country-year data points, with all Gini coefficients happening to refer to disposable income.

To investigate sensitivity of findings to choice of source for the Gini coefficients and knock-on effects in terms of different estimation samples, I ran OLS regressions, with and without dummy variable adjustments, of WIID Reported Gini coefficients on the inflation rate, the unemployment rate, and a time trend, and then examined what happened to the model estimates if the WIID Gini coefficients were substituted by estimates from the LIS Key Figures or Eurostat sources discussed earlier. WIID, LIS, and Eurostat regressions are run with and without common estimation samples. The various regression estimates are shown in Table 4.

Regression 1 provides a 'naïve' reference point: the 727 observations are simply pooled ignoring the multiplicity of observations for many country-year cells and also potential correlations between errors across time within countries. The coefficients on inflation, unemployment, and the time trend are all positive, but not statistically significant in the case of inflation. When the multiple observations per cell are dropped (regression 2), the estimated coefficients change markedly. The coefficient on inflation doubles in magnitude and becomes statistically significant. Similarly the time trend coefficient becomes substantially larger and much more precisely estimated. There is a substantial improvement in goodness of fit: the R^2 increases from 0.028 to 0.145. When country-level cluster-robust standard errors are used (regression 3), the statistical significance of the parameter estimates falls as expected, but they remain statistically significant. Regression 4 implements dummy variable adjustments, by adding regressors that identify differences across observations in equivalence scales (seven types) and in sharing units (household versus family). This adjustment improves overall goodness of fit markedly but also has an impact on the parameter estimates. The coefficients on both inflation and unemployment fall in magnitude and are no longer statistically insignificant. The coefficient on the time trend increases in magnitude, but is less precisely estimated.

When the LIS Key Figures Gini coefficients are substituted for the WIID ones (and observations are dropped if the former are unavailable), the number of observations falls to only 65 for 18 countries (regression 5). The pattern of estimates changes once more. The coefficients on inflation and the time trend are much smaller than in regressions 2–4, and the coefficient on unemployment is larger. All are now statistically significant. Obviously

Table 4 The impacts of unemployment and inflation on income inequality: WIID, LIS, and Eurostat data for Gini coefficients

	Data source for Gini coefficient							
	WIID				LIS	WIID	Eurostat	WIID
Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflation	0.135	0.253	0.253	0.201	0.144	0.123	1.483	1.398
	(0.069)	(0.081)**	(0.078)**	(0.098)	(0.063)*	(0.090)	(0.354)*	(0.304)***
Unemployment	0.257	0.361	0.361	0.281	0.412	0.393	0.251	0.305
	(0.063)***	(0.079)***	(0.152)*	(0.161)	(0.157)*	(0.141)*	(0.160)	(0.155)
Time trend	0.087	0.229	0.229	0.313	0.163	0.283	0.034	0.047
	(0.042)*	(0.047)***	(0.064)**	(0.115)*	(0.073)*	(0.111)*	(0.090)	(0.117)
Intercept	28.536	23.699	23.700	25.437	23.789	24.496	23.421	29.379
	(0.698)***	(0.885)***	(1.660)***	(1.632)***	(1.918)***	(1.480)***	(2.391)***	(1.664)***
DV adjustment?	no	no	no	yes	n. a.	yes	n. a.	yes
R^2	0.028	0.145	0.145	0.374	0.171	0.559	0.246	0.348
Adjusted R^2	0.024	0.134	0.134	0.347	0.130	0.487	0.230	0.319
N (total)	727	242	242	242	65	65	143	143
N (countries)	21	21	21	21	18	18	16	16
First data year	1980	1980	1980	1980	1981	1981	1995	1995
Last data year	2006	2006	2006	2006	2005	2005	2006	2006

OLS estimates, with country-level cluster-robust standard errors for regressions 3–8. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Inflation is the annual change in CPI (%) and unemployment is the fraction of total labour force unemployed (%), with both series taken from (International Monetary Fund (IMF)) (2013). The time trend variable is observation year minus 1990. All Gini coefficients refer to estimates for distributions of disposable income among individuals (with different specific definitions). In regressions 5 and 7, the WIID Gini is replaced by LIS and Eurostat estimates respectively (with observations dropped if no replacement estimate is available). Regressions 6 and 8 use WIID Ginis but the same estimation samples as regressions 5 and 7 respectively. The WIID Gini is the Reported Gini, and refers to distributions with differently defined sharing units and equivalence scales. DV adjustment in regressions 4, 6, 8: regression includes dummy variables identifying differences in equivalence scale (seven types) and differences in sharing unit (household versus family). n. a. not applicable. The LIS Gini estimates refer to incomes defined using the household as sharing unit and square root of household size equivalence scale (source: LIS 2014). The Eurostat Ginis estimates refer to incomes defined using the household as sharing unit and modified-OECD equivalence scale (source: Eurostat 2014, series ilc_di12). Only 'high quality' WIID country-year observations are used. Regression 1 uses all observations in the high-quality subset, regardless of number of 64 observations per country-year cell; all other regressions use data with only one observation per country-year. The 21 countries in regressions (1)–(4) are: AU, AT, BE, CA, DE, EE, ES, FI, FR, GB, GR, IE, IS, IT, LU, NL, NO, NZ, PT, SE, US. See text for further details some of the differences arise from having a different estimation sample. This is illustrated by regression 6 which reverts to using the WIID Gini as the dependent variable but uses the same estimation sample as for the LIS Ginis (as in regression 5) and retains dummy variable adjustments. Comparing the estimates from regression 6 with regression 5, we see that the coefficient on inflation is much the same (0.12 versus 0.14) but it is no longer statistically significant. The coefficient on unemployment is very similar (0.39 versus 0.41) and statistically significant in both cases. However, the coefficient on the time trend is substantially larger (0.28 versus 0.16) and statistically significant.

Regressions 7 and 8 repeat this exercise but, instead, substitute the Eurostat Gini coefficients for the WIID ones or use the WIID Ginis with the regression 7 estimation sample. Now the sample of countries is more homogeneous – by construction, all are EU member states – and the sample is much larger than when the LIS Ginis are used. By contrast with WIID regression 4, the coefficient on inflation in regression 7 is very large (and statistically significant), whereas the coefficients on unemployment and the time trend are smaller and statistically insignificant. The regression 8 estimates show that these contrasts largely arise from the change in estimation sample. Comparison of regressions 7 and 8 shows that corresponding coefficients are similar and so too is their precision.

In sum, these regressions demonstrate that analysts need to exercise care in selecting their estimation samples and to explain and justify their choices. Comparisons of WIID regressions 4 and 8 show that a regression using all observations provides very different results than does a regression based on a homogeneous set of countries (EU member states in this case). The choice of data source for the Gini coefficient also makes a difference but its impact appears to be less marked. That is, whether one uses Eurostat Ginis or WIID Ginis (combined with dummy variable adjustment) can lead to broadly similar estimates – as long as the same estimation sample is used, and this inevitably means a smaller set of countries. But, if this is so, one might ask: why use Gini data sourced from WIID rather than from Eurostat given that the latter are harmonized to a greater extent?

The answer is presumably that one wants estimates for a broader set of countries. The justification for this hinges on the research question, and analysts still need to select their samples carefully for the reasons discussed earlier, to report their selection algorithms, and to justify them. Put another way, a case might be made for using the countries forming the estimation sample for regression 4, but it should be remembered that selection of that sample requires careful case-by-case examination of country data series. A knock-on consequence of the selection exercise was to reduce the number of types of definitional differences, so arguably only more minor ones remain.

Researchers wishing to use wider samples of countries, including 'non-advanced' countries in particular, must address even larger data quality issues than those illustrated here. As Atkinson and Brandolini put it, '[o]ne has to *look at the data*, exercising judgment as to whether they are fit for purpose. Data quality *does* matter' (2009: 399, emphasis in original). Another – complementary – approach is to use dummy variable adjustments, but it is likely that these need to be more sophisticated than simple intercept shifts which assume that definitional differences lead to differences in Gini coefficients that are constant and common across observations (as in the illustration above). As my discussion of the SWIID's construction points out, such assumptions are implausible, and regressions with adjustments should use a number of carefully-defined interaction variables to account for variations in Gini differences across time and space.

Any serious user of WIID must therefore take data quality issues series seriously and, related, they must invest time in understanding original national data sources. To some researchers, these may be unattractive activities. This brings us to the SWIID, for this database offers the potential for avoiding them.

5.2 SWIID-based regressions of the relationship between income inequality, unemployment, and inflation

I now turn to consider regression analysis of the relationship between income inequality, unemployment and inflation using SWIID data. A particular issue is the extent to which results differ from those for WIID and, if so, why. Moreover, there is the additional complication arising from the use of multiply-imputed data – what impact does the uncertainty introduced by this procedure have? Here I examine inequality trends using regression analysis for a range of countries including some non-rich ones.

To begin with, I use the same WEO data as before, so the analysis is restricted to 'advanced economies' again but the period covered now extends to 2012 for some countries (rather than 2006). I fit two sets of OLS regressions, one taking account of imputation variability when calculating standard errors (using Rubin's Rules discussed earlier) and the other ignoring it. Parameter point estimates are the same in each set of regressions.

Column 1 of Table 5 refers to the estimates derived when using all 885 SWIID country-year observations in the selected sample. These refer to 31 countries. The sample size is considerably larger than if all WIID Ginis are used (cf. regressions 2–4 in Table 4). In this regression 1, the coefficients on inflation and unemployment are not statistically significant; only that on the time trend and for the intercept are. The regression fit in terms of R^2 is much lower than for the WIID (Table 4, regressions 3, 6, 8), perhaps reflecting the greater diversity of countries included in SWIID estimation sample. Interestingly, although the MI-estimated standard errors are larger than their 'ordinary' counterparts, the difference is negligible in this case, and this is also true for all regressions shown in the table.

Regression 2 restricts the estimation sample to countries that are OECD member states and joined before 1990 (23 countries). The estimated parameters and model fit change markedly. Although the coefficient on inflation remains insignificant, its magnitude increases substantially. The coefficient on unemployment doubles, to 0.34, and become statistically significant. The intercept increases from 0.09 to 0.15 and is more precisely estimated. Regression 3 restricts attention to an even more homogeneous sample, the EU-15. Again the parameter estimates change. The coefficient on inflation remains insignificant but increases in magnitude; the coefficient on unemployment increases in magnitude to 0.43, is statistically significant but more precisely estimated. The same is true for the time trend. Adjusted R^2 increases from 0.157 to 0.255.

In regression 4, the estimation sample is restricted further, to EU15 member states and data for 1995–2006, so corresponds closely to the sample used in regressions 7 and 8 in Table 4 (using Eurostat Ginis and WIID Ginis with dummy variable adjustment, respectively). Restricting the estimation period has a big impact on the estimates (compare SWIID regression 4 with regression 3) and the result is a much closer correspondence between the estimates using WIID and the other data sources: the coefficient on inflation is relatively large and statistically significant and the coefficients on unemployment and the time trend are not statistically significant: compare regression 4 in Table 5 with regressions 7 or 8 in Table 4.

Overall, the SWIID regressions suggest some conclusions that are similar to the WIID ones: estimation results are sensitive to the choice of sample, both in terms of country

Table 5 Multiple imputation estimates of the impacts of unemployment and inflation on income inequality: SWIID data for net income Gini coefficients

	Estimation subsample			
	All observations	Pre-1990 OECD members	EU15	EU15 (1995–2006)
Regressor	(1)	(2)	(3)	(4)
Inflation	0.012	0.143	0.249	1.517
	(0.007)	(0.112)	(0.135)	(0.349) ***
	(0.007)	*(0.110)*	*(0.133)*	*(0.347)* ***
Unemployment	0.166	0.335	0.428	0.333
	(0.180)	(0.101) **	(0.084) ***	(0.184) *
	(0.179)	*(0.101)* **	*(0.083)* ***	*(0.184)* *
Time trend	0.090	0.146	0.191	0.119
	(0.032) **	(0.037) ***	(0.038) ***	(0.071)
	(0.032) **	*(0.037)* ***	*(0.038)* ***	*(0.070)*
Intercept	27.108	24.647	22.491	21.716
	(1.912) ***	(1.427) ***	(1.472) ***	(2.513) ***
	(1.906) ***	*(1.422)* ***	*(1.465)* ***	*(2.503)* ***
R^2	0.044	0.160	0.259	0.264
Adjusted R^2	0.040	0.157	0.255	0.251
N (total)	885	715	472	180
N (countries)	31	23	15	15

OLS estimates, with country-level cluster robust standard errors. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. The SE estimates in parentheses are multiple imputation estimates, accounting for imputation variability. The italicized SE estimates in parentheses do not take account of imputation variability. The R^2 and Adjusted R^2 statistics refer to OLS regressions that do not account for imputation variability. In every regression, the data cover the period 1980–2012 with the exception of regression 4, in which case the period is restricted to 1995–2006. All observations: all country-year observations from 'advanced economies' with non-missing inflation and unemployment data from International Monetary Fund (IMF) (2013) and non-missing SWIID data on the net income Gini coefficients

and year coverage. If the estimation sample is restricted to a homogenous set of countries (EU15), then it appears that SWIID and WIID provide similar estimates. At the other extreme, if one mechanically fits regressions to estimation samples that maximize country and period coverage, then WIID and SWIID provide very different results concerning the relationship between income inequality, unemployment and inflation.

Bias in SWIID regression point estimates derived from samples with global coverage cannot be assessed because there are no relevant external benchmarks (see earlier). However, interestingly, the SWIID regressions suggest that properly accounting for imputation variability increases standard error estimates only marginally and does not change conclusions about statistical significance. This is reassuring, if only because previous users of SWIID data appear to have ignored the multiply-imputed nature of their data. See, for example, Acemoglu et al. (2015), Ostry et al. (2014), and Solt (2011).

This finding needs to be checked further, especially for situations in which estimation samples are extended to include developing countries for which data quality is lower and imputation variability is much greater. This motivates the next set of regressions.

5.3 SWIID-based regressions of inequality trends

Table 6 shows the estimates derived from country-specific OLS regressions of SWIID Gini coefficients on an intercept and binary indicator variables for the four decades from 1970 onwards. The intercept is the country's Gini coefficient for the decade 1960–69 (on average; in per cent), and the coefficient on each decade indicator shows the difference between inequality in that decade and 1960–69. The six countries represent the full range of SWIID imputation variability, from the UK and USA with relatively little variability to China and Kenya with a lot. As in Table 5, standard errors are presented both taking into account the imputation variability and ignoring it. Point estimates are the same in both cases.

The table suggests that ignoring imputation variability makes little difference to inference about inequality trends, except in cases in which variability is extremely high. This is shown in the top right-hand side of the table where, for example the coefficients for 1970–79 for China and Kenya differ insignificantly from zero when MI estimation methods are used but are statistically significant when they are not.

Other regressions (not shown) support the conclusion about the impact of imputation variability depending on the prevalence of high variability. If the regressions are repeated

Table 6 Multiple imputation estimates of inequality trends: SWIID data for net income Gini coefficients, by country

Regressor	UK	USA	Finland	Argentina	China	Kenya
Decade: 1970–1979	–0.056	3.032	–5.882	–0.134	–3.434	–7.756
	(0.556)	(0.551) ***	(0.923) ***	(1.532)	(3.928)	(8.043)
	(0.474)	*(0.446)* ***	*(0.667)* ***	*(1.327)*	*(1.993)* *	*(2.783)* ***
Decade: 1980–1989	2.111	4.591	–10.887	3.210	–7.471	–9.130
	(0.550) ***	(0.550) ***	(0.868) ***	(1.519) **	(3.646) *	(10.433)
	(0.474) ***	*(0.446)* ***	*(0.667)* ***	*(1.327)* **	*(1.938)* ***	*(2.517)* ***
Decade: 1990–1999	7.380	8.014	–9.544	6.386	6.306	–20.025
	(0.542) ***	(0.558) ***	(0.873) ***	(1.492) ***	(3.684) ***	(7.095) **
	(0.474) ***	*(0.446)* ***	*(0.667)* ***	*(1.327)* ***	*(1.938)* ***	*(2.517)* **
Decade: 2000–2012	8.217	9.594	–5.918	6.218	14.704	–22.501
	(0.520) ***	(0.536) ***	(0.843) ***	(1.452) ***	(3.698) ***	(7.991) **
	(0.446) ***	*(0.427)* ***	*(0.627)* ***	*(1.283)* ***	*(1.938)* ***	*(2.783)* ***
Intercept	26.722	27.572	31.460	36.976	34.638	67.827
	(0.412) ***	(0.440) ***	(0.722) ***	(1.272) ***	(3.541) ***	(6.658) ***
	(0.335) ***	*(0.316)* ***	0.471 ***	*(1.122)* ***	*(1.700)* ***	*(2.055)* ***
R^2	0.929	0.931	0.870	0.629	0.904	0.764
Adjusted R^2	0.923	0.925	0.859	0.594	0.893	0.735
N (years per country)	53	52	53	47	41	37

OLS estimates. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. The SE estimates in parentheses are multiple imputation estimates, accounting for imputation variability. The italicized SE estimates in parentheses do not take account of imputation variability. The R^2 and Adjusted R^2 statistics refer to OLS regressions that do not account for imputation variability. The Intercept estimates show the net income Gini coefficients for '1960–1969' (in per cent). The coefficient estimate for each other decade shows the change in the Gini between the decade in question and the decade '1960–1969' (the latter is the omitted category for the Decade classification) using samples that include all countries from a region rather than a single country, the differences between MI estimates and non-MI estimates are smaller (not shown). For example, with a sample for Africa, the extreme imputation variability for Kenya plays less of a role than it does if one looks at Kenya alone.

Using the same data set as used in the previous subsection, I have also explored whether the conclusions about the impact on estimates of imputation variability carry across to models in which the Gini coefficient is an explanatory variable rather than the dependent variable, and to non-linear models (a Poisson regression, specifically). In these cases, properly accounting for MI led to changes in point estimates relative to those for regressions that ignore MI but these changes are small and, again, the changes in standard errors are also relatively small. Of necessity, these regressions pool data from multiple countries, and so there is less chance that extreme imputation variability has a big impact.

6 Conclusions

WIID and SWIID are resources with substantial potential, but there are also pitfalls facing users. The problems that Atkinson and Brandolini (2001, 2009) drew attention to with reference to the predecessors of WIID remain. Researchers must take care when selecting observations, confront the very real data quality issues head-on, and check whether their conclusions are robust to different treatments of the data. The advent of SWIID raises new issues about the quality of the imputations per se and about how to account for multiply-imputed observations in estimation.

Researchers also need to confront an uncomfortable and inevitable trade-off between country coverage and data quality. A focus on a relatively small number of homogeneous countries such as OECD or EU member states is accompanied by availability of data of higher quality. (Indeed, if country groups such as these are the interest, or single-country studies, then sources other than WIID and SWIID should be used.) Broadening the scope of analysis to take a more global perspective inevitably means that the secondary data on inequality are of poorer quality, as represented by a lower quality assessment in WIID. There is also a higher prevalence of missing data, and hence a greater proportion of the observations in SWIID more heavily reliant for their accuracy on the validity of the imputation model, and there is greater imputation variability. There is inevitably a degree of uncertainty associated with estimates derived from samples with global coverage, whether based on WIID or SWIID, and it is different from the uncertainty arising from sampling variability. It is a type of systematic measurement error that is not 'classical' in form (arguably the magnitude of the error is correlated with the true value of the Gini, for example).

Clearly, WIID and SWIID offer researchers very different strategies for handling issues of missing and non-comparable data. My analysis leads me to be more sympathetic to an approach that works directly with the data points in WIID and benchmarks them against national sources wherever possible, rather than taking advantage of the convenience that the SWIID offers. Put another way, I believe empirical researchers should take responsibility for checking the data with which they work and its quality. Also, since the relevance of different types of data non-comparability and quality are likely to be specific to the research question considered, a universal problem-solving approach as provided by SWIID is less desirable. Important details of the SWIID approach are hidden from the user in any case. Moreover, there are questions concerning the imputation model that underpins SWIID. To me, SWIID provides plausible data but not sufficiently credible data.

My analysis suggests that the principal issues with using SWIID data concern potential bias rather than precision. If the imputation model is problematic, the data provided about inequality levels and trends within and between countries – and their relationships with other variables – are incorrect, and regression point estimates will be wrong. (The magnitude of this bias is difficult to assess because, by definition, there are no external benchmarks for all observations in samples with global coverage.) In contrast, the illustrative regression analysis suggests that ignoring imputation variability and simply using imputation averages may not lead to standard errors that are too far wrong – though this conclusion is conditional on the sample employed (and inference overall also depends on the point estimates being right). The more countries with high imputation variability there are in an estimation sample, the greater the risk of incorrectly finding statistically significant results. Since MI estimation procedures are widely available in statistical software nowadays, and for many non-linear models as well as linear ones, researchers should employ these methods to inoculate themselves against this risk.

In sum, from a data issues perspective, I recommend WIID over SWIID, and my support for researchers' use of WIID is conditional. At the very least, WIID-based papers should report and justify the algorithm that the authors used to select their sample, including the selection rules applied to situations in which there are multiple observations per country-year cell. Checks on the robustness of findings to different selection algorithms are also important. In addition, regression-based adjustment procedures to account for differences in definitions need to go beyond use of simple intercept shifts that implement assumptions of constant and common differences across observations. More flexibility can be introduced using judicious interactions between key explanatory variables and country/region and time.

Since there are clearly potential costs arising with the use of any world income inequality database, researchers also need to spell out the benefits of their chosen strategy in order to convince readers that the benefit-cost ratio is favourable.

Some specific questions for further research include the following. Within which groupings of country-year observations is it sufficiently plausible to assume that Gini differences (as in the dummy variable adjustment approach in regressions using WIID) or Gini ratios (as in the SWIID imputation model) are constant? And is it better to work with differences or ratios? Which types of non-comparability are the most important, and in the context of which types of research question?

There is enough evidence already available to raise questions about assumptions of constancy of differences or ratios within broadly-defined groups of observations. In addition, there are tricky questions concerning how to utilize information about the heterogeneous quality of individual country-year observations in estimation and imputation. Systematic assessment of these issues is required in order to inform the use of both WIID and SWIID.

Acknowledgments The research for this paper was partially supported by core funding of the Research Centre on Micro-Social Change at the Institute for Social and Economic Research by the University of Essex and the UK Economic and Social Research Council (award RES-518-28-001). My thanks go to Facundo Alvaredo, Tony Atkinson, Andrea Brandolini, Chico Ferreira, Markus Jäntti, Stephan Klasen, Christoph Lakner, Nora Lustig, Branko Milanovic, Tim Smeeding, and Philippe Van Kerm for helpful comments and conversations. For feedback on a first draft, I am grateful to Nina Badgaiyan, Jukka Pirttilä, and Finn Tarp (WIID) and Frederick Solt (SWIID). I have also drawn on documents summarising the WIID and SWIID and research based on them that were prepared by Nicole Florack and Dan Teles. All data analysis was undertaken using Stata version 13.1. Replication do-file code and data are available from the author on request.

References

Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J.A.: Democracy, redistribution and inequality. In: Atkinson, A.B., Bourguignon, F. (eds.) Handbook of Income Distribution, vol 2. Elsevier, North-Holland (2015)

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E.: The World Top Incomes Database, online database. http://topincomes.g-mond.parisschoolofeconomics.eu/ (2014). Accessed 4 Apr 2014

Atkinson, A.B., Brandolini, A.: Promise and pitfalls in the use of "secondary" data-sets: income inequality in OECD countries as a case study. J. Econ. Lit. **34**(3), 771–799 (2001)

Atkinson, A.B., Brandolini, A.: On data: a case study of the evolution of income inequality across time and across countries. Camb. J. Econ. **33**(3), 381–404 (2009)

Atkinson, A.B., Rainwater, L., Smeeding, T.M.: Income distribution in OECD countries: evidence from the Luxembourg Income Study, OECD Social Policy Studies No. 18. OECD, Paris (1995)

Barro, R.J.: Inequality and growth in a panel of countries. J. Econ. Growth **5**(1), 5–32 (2000)

Barro, R.J.: Inequality and growth revisited', Working Paper Series on Regional Economic Integration No. 11. Manila: Asian Development Bank. http://aric.adb.org/pdf/workingpaper/WP11_%20Inequality_and_Growth_Revisited.pdf (2008). Accessed 30 Mar 2014

Berg, A., Ostry, J.D., Zettelmeyer, J.: What makes growth sustained? J. Dev. Econ. **98**(2), 149–166 (2012)

Blinder, A., Esaki, H.: Macroeconomic activity and income distribution in the postwar United States. Rev. Econ. Stat. **60**(4), 604–609 (1978)

Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J.: Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring. J. Econ. Inequal. **9**(3), 393–415 (2011)

Castelló-Climent, A.: Inequality and growth in advanced economies: an empirical investigation. J. Econ. Inequal. **8**(3), 293–321 (2010)

Chambers, D., Krause, A.: Is the relationship between inequality and growth affected by physical and human accumulation? J. Econ. Inequal. **8**(2), 153–172 (2010)

Chotikapanich, D., Griffiths, W.E., Rao, D.S.P., Valencia, V.: Global income distributions and inequality, 1993 and 2000: incorporating country-level inequality modeled with Beta distributions. Rev. Econ. Stat. **94**(1), 52–73 (2012)

Clark, R.: Convergence in national income distributions. Social Forces **92**(2), 413–436 (2013)

Coulter, F.A., Cowell, F.A., Jenkins, S.P.: Equivalence scale relativities and the extent of inequality and poverty. Econ. J. **102**(414), 1067–1082 (1992)

Deninger, K., Squire, L.: A new dataset measuring income inequality. World Bank Econ. Rev. **10**(3), 565–591 (1996). Data available from http://go.worldbank.org/UVPO9KSJJ0. Accessed 30 Mar 2014

Eurostat: Statistics, online database. http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database (2014). Accessed 6 Feb 2014

Forbes, K.J.: A reassessment of the relationship between inequality and growth. Am. Econ. Rev. **90**(4), 869–887 (2000)

Gruen, C., Klasen, S.: Growth, inequality, and welfare: comparisons across space and time. Oxf. Econ. Pap. **60**(2), 212–236 (2008)

Gruen, C., Klasen, S.: Has transition improved well-being? Econ. Syst. **36**(1), 11–30 (2012)

International Monetary Fund (IMF): World Economic Outlook, April 2013 edition online database. http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/index.aspx (2013). Accessed 23 Mar 2014

Institute for Fiscal Studies (IFS): Inequality and Poverty Spreadsheet (to accompany IFS Report on Living Standards, Poverty and Inequality in the UK: 2013, file bn19figs.xlsx. http://www.ifs.org.uk/fiscalFacts/povertyStats (2013). Accessed 8 Oct 2013

Jäntti, M., Jenkins, S.P.: Examining the impact of macro-economic conditions on income inequality. J. Econ. Inequal. **8**(2), 221–240 (2010)

Layte, R.: The association between income inequality and mental health: testing status anxiety, social capital, and neo-materialist explanations. Eur. Sociol. Rev. **28**(4), 498–511 (2012)

LIS Cross-National Data Center (LIS): Key Figures, online database. http://www.lisdatacenter.org/data-access/key-figures/ (2014). Accessed 6 Feb 2014

Organisation for Economic Cooperation and Development (OECD): Divided We Stand. Why Inequality Keeps Rising. OECD, Paris (2011)

Ostry, J.D., Berg, A., Tsangarides, C.G.: Redistribution, inequality, and growth, IMF Staff Discussion Note SDN1402. International Monetary Fund, Washington. http://www.imf.org/external/pubs/ft/sdn/2014/sdn1402.pdf (2014). Accessed 17 Mar 2014

Parker, S.C.: Income inequality and the business cycle: a survey of the evidence and some new results. J. Post-Keynesian Econ. **21**(2), 201–225 (1998)
Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., Solenberger, P.: A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv. Methodol. **27**(1), 85–95 (2001)
Rubin, D.B.: Multiple Imputation for Nonresponse in Surveys. Wiley, New York (1987)
Rubin, D.B.: Multiple imputation after 18+ years. J. Am. Stat. Assoc. **91**(434), 473–489 (1996)
Sala-i-Martin, X.: The world distribution of income: falling poverty and ... convergence, period. Q. J. Econ. **121**(2), 351–397 (2006)
Shorrocks, A., Wan, G.: Ungrouping income distributions. Synthesising samples for inequality and poverty analysis. Research Paper No. 2008/16. Helsinki: UNU-WIDER. http://www.wider.unu.edu/publications/working-papers/research-papers/2008/en_GB/rp2008-16/ (2008). Accessed 31 Jan 2014
Solt, F.: Standardizing the world income inequality database. Soc. Sci. Q. **90**(2), 231–242 (2009)
Solt, F.: Diversionary nationalism: economic inequality and the formation of national pride. J. Polit. **73**(3), 821–830 (2011)
Solt, F.: SWIID. The Standardized World Income Inequality Database, version 4.0 September 2013. University of Iowa, Iowa City. http://myweb.uiowa.edu/fsolt/swiid/swiid.html (2013a). Accessed 31 Jan 2014
Solt, F.: The Standardized World Income Inequality Database', Presentation to the Development Policy and Analysis Division, Department of Economic and Social Affairs, United Nations, New York, 18 September 2013, http://myweb.uiowa.edu/fsolt/papers/SWIID%20UN%20Talk.pdf (2013b). Accessed 18 Aug 2014
Teulings, C., van Rens, T.: Education, growth, and income inequality. Rev. Econ. Stat. **90**(1), 89–104 (2008)
UNU-WIDER: UNU-WIDER World Income Inequality Database, Version 2.0c, May 2008. UNU-WIDER, Helsinki. http://www.wider.unu.edu/research/Database/en_GB/database/ (2008). Accessed 31 Jan 2014
Xie, Y., Zhou, X.: Income inequality in today's China. Proc. Natl. Acad. Sci. U.S.A. **111**(19), 6928–6933 (2014)