Economics Bulletin

Volume 38, Issue 4

A note on the use of decile or quintile group-share of income or consumption from the popular income inequality databases to explain inequality conditions

Amlan Majumder
University of North Bengal, Darjeeling

Takayoshi Kusago
Kansai University, Osaka

Abstract

As Gini coefficient does not summarise all that an income distribution can tell us about inequality, there has been an increasing thrust in literature on supplementing or even replacing the use of it by direct examination of the income distributions. However, the problem that remains with the common readers is that they are to rely on popular databases of inequality for Gini coefficient based on microdata; and on income distributions, which are squeezed into deciles or quintiles. The basic question is that whether such grouped data are consistent enough to do the practices as stated above. Any doubtful use of those may be very misleading. In such a situation, we perform some consistency checks of inequality data available in various World Development Indicators and World Income Inequality Database - WIID 3.4 for illustrative purpose only in interest of the common readers. In this connection, we discuss about the issues like shortfall, underestimation, bias etc. As observed from the preliminary results, nearly 7.5 % and 13 % cases appear to be unusual in quintile and decile data respectively in WIID 3.4. Some of these are simply misreporting or typos, which are to be corrected, and some appear to be very special (with positive bias instead of downward bias), which warrant theoretical attention for further research. Common readers may restrain themselves from using the unusual cases in study to explain inequality conditions.

We are grateful to the anonymous referee and Stéphane Mussard, Associate Editor, Economics Bulletin for valuable comments and suggestions. Responsibility for any error rests with us.
Citation: Amlan Majumder and Takayoshi Kusago, (2018) "A note on the use of decile or quintile group-share of income or consumption from the popular income inequality databases to explain inequality conditions", *Economics Bulletin*, Volume 38, Issue 4, pages 2152-2166
Contact: Amlan Majumder - amlan@amlan.co.in, Takayoshi Kusago - tkusago@kansai-u.ac.jp.
Submitted: September 05, 2018. **Published:** December 02, 2018.

Economics Bulletin

Submission Number: EB-18-00718

A note on the use of decile or quintile group-share of income or consumption from the popular income inequality databases to explain inequality conditions

Amlan Majumder
Dinhata College, West Bengal, India

Takayoshi Kusago
Kansai University, Osaka

Abstract

As Gini coefficient does not summarise all that an income distribution can tell us about inequality, there has been an increasing thrust in literature on supplementing or even replacing the use of it by direct examination of the income distributions. However, the problem that lies with the common readers is that they are to rely on popular databases of inequality for Gini coefficient based on microdata; and on income distributions, which are squeezed into deciles or quintiles. The basic question is that whether such grouped data is consistent enough to do the practices as stated above. Any doubtful use of those may be very misleading. In such a situation, we perform some consistency checks of inequality data available in various World Development Indicators and World Income Inequality Database - WIID 3.4 for illustrative purpose only in interest of the common readers. In this connection, we discuss about the issues like shortfall, underestimation, bias etc. As observed from the preliminary results, nearly 7.5 % and 13 % cases appear to be unusual in quintile and decile data respectively in WIID 3.4. Some of these are simply misreporting or typos, which are to be corrected, and some appears to be very special (with positive bias instead of downward bias), which warrant theoretical attention for further research. Common readers may restrain themselves from using the unusual cases in study to explain inequality conditions.

Submitted: September 05, 2018.

1. Introduction

Gini coefficient does not summarise all that an income distribution can tell us about inequality. As a result, there has been an increasing thrust in literature on supplementing or even replacing the use of the said measure by direct examination of the income distributions. As examination of income distributions based on microdata (at country level) is not always possible for common readers, they are to rely on popular databases of inequality for: (i) Gini coefficient based on microdata (micro-Gini henceforth), and (ii) income distributions, which are squeezed into deciles or quintiles. The basic question at this point is that whether such grouped data are consistent enough to do the said practices as thrusted in literature. Even if we ignore misreporting and typos, there may remain issues with the process of squeezing of microdata. However, if any such issue remains, the use of decile and quintile group-share of income may be very misleading to explain inequality conditions. The issues at the second stage are that whether and about how we can check consistency of grouped data available in popular databases. These are some of the pertinent questions, which this paper tries to address for illustrative purpose only utilising data from the World Income Inequality Database - WIID 3.4 (UNU-WIDER, 2017) and various World Development Indicators.

Although we are habituated to use grouped income distributions to explain inequality conditions for long, the need for the same has been reiterated recently by the leading authors in the field. For example, Osberg (2017) uses his classical illustration of ‘Adanac’ (Osberg, 1981, p. 14) recently to justify the need for supplementing the use of a summary measure by direct examination of the income distribution. ‘Adanac’ considers a simple two-class example in which the Gini coefficient is held constant while the size of the rich and poor changes implying a series of different income distributions, which represent a constant Gini coefficient. In such cases in reality, one must rely on visual examination of the income distributions. In addition to the above, also there are references in literature on straightforward replacement of the use of Gini coefficient as a summary measure by visual examination of income shares from distribution tables (see Piketty, 2014, p. 266). As the need for using income distributions to explain inequality conditions is stressed and reiterated in literature, the selection of the topic of research under discussion is too relevant in the present context.

2. Methods of checking consistency of data in squeezed income distributions

It is to be noted that there is no hard and fast rule to check consistency of decile or quintile group-share of income or consumption. Following a trial and error procedure, we have found two ways, which can be adopted separately or jointly to perform consistency check of grouped data. Under the first method, we compute Gini coefficients from the decile or quintile income distributions. Such Gini coefficients may be termed as decile-Gini or quintile-Gini henceforth. We then draw a scatterplot of micro-Gini (along vertical axis) and decile- or quintile-Gini (along horizontal axis) and fit a regression line. From our repeated exercises, we have seen that if the data quality is perfect in both the sides (i.e., the calculation and presentation of micro-Gini as well as those associated with decile or quintile group-share of income), the scatterplot will show a perfect linear relationship between the two with an R-square value of 1.00. If there exists any discrepancy in the data, it will be indicated by distortions and lower goodness of fit of the said relationship. In such situations, readers may restrain themselves from using the distorted observations in study.

In order to proceed with the second method, we need to be aware that Gini coefficient exhibits a shortfall in its value due to grouping of observations into smaller number of parts. It implies that Gini coefficient, when computed from microdata (usually with n > 5000), turns out to be larger in comparison to that computed from the same data squeezed into say, ventile (for k = 20) or decile (for k = 10) or quintile (for k = 5), where k = number of equally sized groups. The shortfall increases as the group-number (k) decreases. Under this method, we compute shortfall as a percentage of micro-Gini and fix cut-off points for k = 5 and 10. If for any country or case, shortfall exceeds the cut-off point, it indicates presence of discrepancy in data. In such cases readers may restrain themselves from supplementing the use of micro-Gini by visual examination of the squeezed distributions.

Fixing cut-off points for shortfall is not easy. Empirical literature on shortfall due to grouping of microdata is very limited. Although intense, theoretical literature too on the subject matter is not vast. However, everybody associated with the measurement techniques of economic inequality might be aware that the loss of information generated under the process of squeezing of microdata has a distribution-free part, which is certain; the remaining portion (if arises) is distribution specific and is uncertain. For the sake of simplicity we may call the distribution-free part as 'underestimation' and the stochastic part as '(downward) bias' of Gini coefficient respectively.

van Ourti and Clarke (2011), in their seminal paper, reported from the study of Lerman and Yitzhaki (1989), that the shortfalls[1] from using grouped data with ten and five income categories are about 2.5 % and 7 % respectively of the Gini coefficients as calculated from microdata. In order to have a 'simple' correction for the distribution-free part of it, van Ourti and Clarke (2011) suggested a correction factor: $\mathrm{k}^2/(\mathrm{k}^2-1)$, where k = number of equally sized groups. From this factor it is evident that magnitude of underestimation due to grouping is: $\{1-(\mathrm{k}^2-1)/\mathrm{k}^2\}$. It implies that if we work with decile group-share of income, where k = 10, the magnitude of underestimation is:$[100 * \{1-(10^2-1)/10^2\}]$ = 1 %. When the total shortfall (for k = 10) is 2.5 % (as cited above), the magnitude of (downward) bias may go up to 1.5 %. Similarly, when k = 5, underestimation will be of 4 % and (downward) bias may go up to 3 % (if the total shortfall is of 7 %, as cited above).

As we have found just one reference on magnitude of shortfall for k = 5 and 10, we cannot fix cut-off points from it. In practice we have seen that for k = 5, the admissible magnitude of shortfall may go slightly beyond 11 % provided that a perfect (or nearly perfect) linear relationship (with an R-square value of 1.00 or close to 1.00) is maintained between micro-Gini and quintile-Gini. For the sake of simplicity, we may fix the cut-off point at 12 % for k = 5. As it contains a fixed amount of underestimation of 4 %, the admissible amount of shortfall in case of five income groups may range from 4 % to 12 %.

Under a similar condition, for k = 10, the admissible magnitude of shortfall may go slightly beyond 4 %. However, for the sake of simplicity, we may fix a cut-off point at 5 % for k = 10. As it contains a fixed amount of underestimation of 1 %, the admissible amount of shortfall in case of ten income groups may range from 1 % to 5 %.

We understand from the above that if we adopt both the methods of checking consistency together, it will make enough ground for us to determine whether there exists any discrepancy in data or whether a simple visual linking of the micro-Gini with the decile or quintile income distributions is possible.

[1] They used the term 'bias' for the whole shortfall, and termed the distribution-free and distribution specific parts of it as 'first order bias' and 'second order bias' respectively.

3. Some examples of checking consistency of microdata and squeezed data

We have one ready reference with us where, Milanovic (2012)[2] computed Gini coefficient from microdata for some countries with large number of observations ($5227 \le n \le 65809$) and then he squeezed microdata for each of the countries into twenty ventiles and computed Gini coefficients again. He then calculated shortfall as a percentage of micro-Gini for each such case. In his exercise (for k = 20), the shortfall ranges from 0.6 % to 1.1 %, which contains a fixed amount of underestimation of 0.25 %, as shown in table 1 below.

Table 1. Example of shortfall of Gini coefficient in some countries

Countries (Year)	g (m)	n	g (v)	k	Shortfall* (%)	Underes-timation (%)	Downward Bias (%)
Belarus (2006)	28.67	5227	28.50	20	0.6	0.25	0.35
Germany (2005)	31.49	11197	31.23	20	0.8	0.25	0.55
Poland (2005)	34.49	34767	34.23	20	0.8	0.25	0.55
Indonesia (2005)	39.41	64595	39.01	20	1.0	0.25	0.75
Bangladesh (2005)	41.23	10080	40.86	20	0.9	0.25	0.65
Iran (2005)	41.92	26850	41.80	20	0.3	0.25	0.05
Uganda (2005)	42.94	7421	42.54	20	0.9	0.25	0.65
Mexico (2004)	45.72	22554	45.56	20	0.3	0.25	0.05
Kenya (2004-05)	47.62	13158	47.10	20	1.1	0.25	0.85
Chile (2003)	54.56	65809	53.96	20	1.1	0.25	0.85

G (m): Gini coefficient computed from microdata; G (v): Gini coefficient computed from ventiles; * Milanovic (2012) used the term 'underestimation' for the whole shortfall.
Source: Table 3 of Milanovic (2012) and self-elaboration

Now, we may check the empirical relationship between micro-Gini and that obtained from grouped data (say ventile Gini). It has been observed (empirically) that, in the absence of anomaly, those are perfectly correlated maintaining a linear relationship with an R-square value of 1.00. For example, we explore the relationship between the two by drawing a scatterplot and fitting a regression line, as shown in figure 1 below.

In reference to the discussion as presented above, it is clear from figure 1 that Milanovic (2012) has been meticulous in doing his exercise - in the process of computing micro-Ginis as well as in the process of squeezing of microdata into ventiles. Although he did not present the squeezed income distributions, if available, one can simply link those to the respective micro-Ginis for visual examination without any hesitation.

We now redo one exercise as originally done by Majumder (2010) on exploring the relationship between micro-Gini and those computed from quintile data available in World Development Indicators 2009. The details of the relationship are presented below in figure 2.

By looking at the figure 2, we realise that the model lags behind of that in figure 1 in terms of goodness of fit, as reflected from the comparatively low R-square value. There are also some noticeable distortions in it. In cases of Denmark and Sweden, after squeezing of microdata, quintile-Ginis supersede the respective micro-Ginis as shown in table 3 below.

From table 3 we see that after squeezing of microdata, Gini coefficients of Denmark and Sweden inflated by 2.43 % and 2.00 % respectively. The same figures (for micro-Gini) were reported in WDI 2009 to WDI 2011 for the two countries. Although not included in figure 2, the

[2] The objective of the study is different from that of the present one.

same thing happened for Togo also in 2009 and 2010. After squeezing of microdata, Gini coefficient for Togo increased by a reasonable amount.

Figure 1. Relationship between Gini coefficients obtained from microdata and ventile data as presented by Milanovic (2012); n= 10

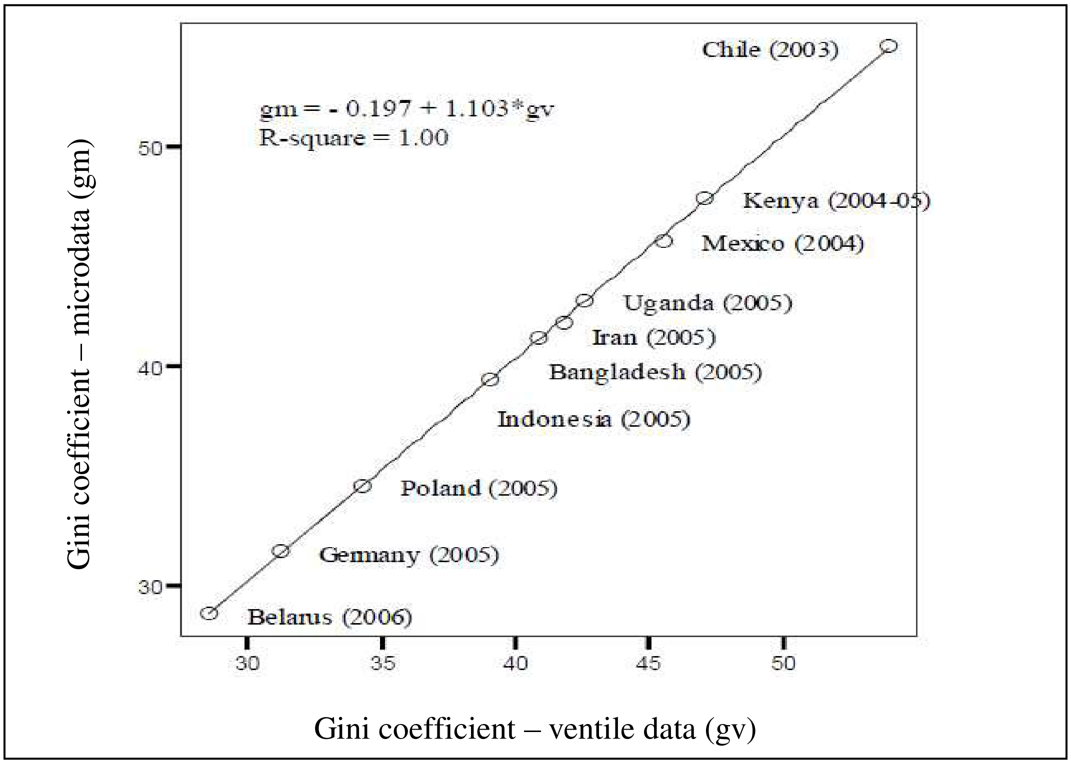


Figure 2. Relationship between Gini coefficients obtained from microdata and quintile data as presented in World Development Indicators 2009; n= 135

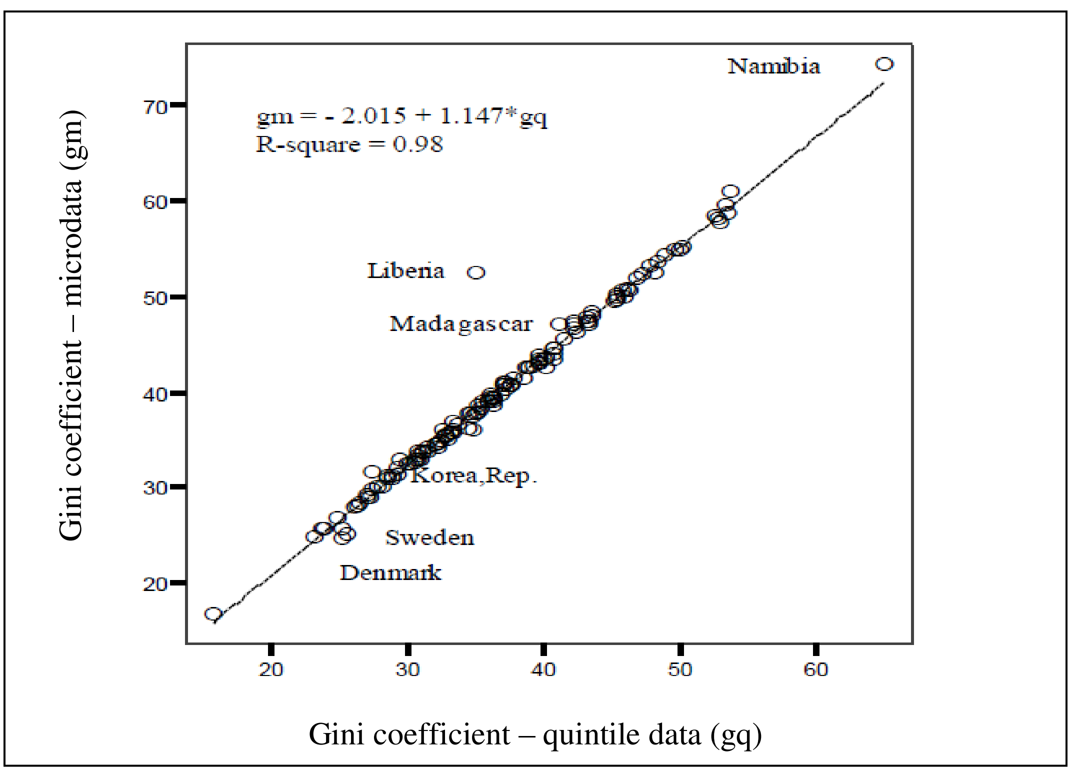


Table 3. Some unusual cases as found in some World Development Indicators (WDIs)

Country	WDI	Gini coefficient (Microdata)	Gini coefficient (Quintile data)	Shortfall	Shortfall (%)	Under-estim-ation (%)	Downw-ard Bias (%)
Denmark	2008-2011	24.70	25.30	-0.60	-2.43	4.00	-6.43
Norway	2008-2012	25.80	25.30	0.50	1.94	4.00	-2.06
Sweden	2008-2011	25.00	25.50	-0.50	-2.00	4.00	-6.00
Togo	2010-2011	34.40	38.00	-3.60	-10.58	4.00	-14.58
UK	2008-2014*	36.00	34.80	1.20	3.33	4.00	-0.67

* Figures in the quintile distribution are rounded off in WDIs 2013 & 2014 leading to slightly different Gini coefficient.

Source: WDIs 2008 to 2014 and self-elaboration

The instances of Norway and United Kingdom are also to be noted. After squeezing of microdata, Gini coefficient in each of these two countries decreased to some extent. However, if we consider the amount of fixed underestimation due to grouping (i.e., 4 %), these two countries too exhibit positive bias of 2.06 % and 0.67 % respectively. If we look at the WDIs from 2008 to 2014[3], there are 18 such cases, where amount of shortfall in each is less than 4 % indicating existence of positive bias ranging from 0.19 % (Singapore in WDI 2013) to 2.05 % (Mexico in WDI 2008).

In figure 2, four other unusual cases are also highlighted (Korea Republic, Liberia, Madagascar and Namibia), where shortfall ranges approximately between 12 % and 33 %. However, not all such cases need theoretical attention. There are some cases, which occurred either due to inappropriate reporting of micro-Gini or probably due to inappropriate process of squeezing of microdata into quintiles. For example, the case of Liberia has simply been a misreporting, as we came to know from the finding of Majumder (2010) and the 'Data Updates and Errata' published by the World Bank in April 2011[4]. The corrected Gini coefficient for Liberia, as reported by the Bank in the said source, is 38.2 (instead of 52.6). Squeezing of the said Liberian data into quintiles generates a shortfall of 8.38 %, which falls within the admissible range.

Except some anomalies, the cases, which are highlighted above, are very special, as literature on existence of positive bias in studies of shortfall of Gini coefficient due to grouping is less known. Researchers should take a note of these special cases for further research.

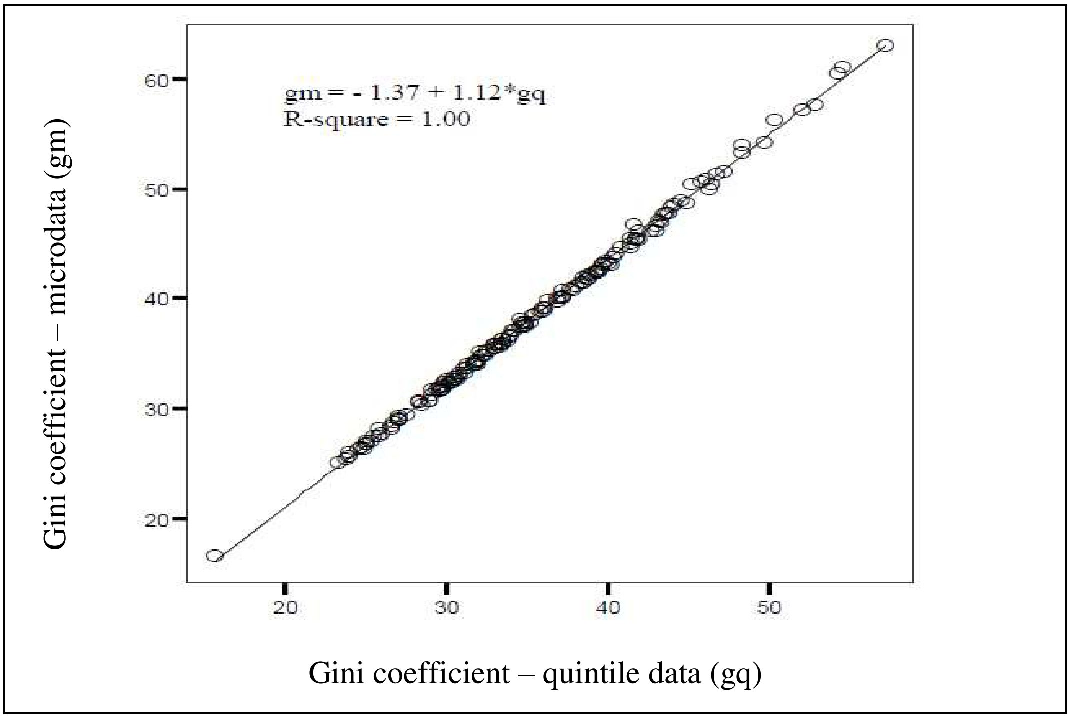
Existence of positive bias also creates distortions. One may check in the figures that all the distorted cases lying towards the right of the regression line indicate positive biases, which are prominent in figures 4 and 7. Distorted cases lying towards the left of the regression line indicate shortfall (underestimation plus downward bias, as applicable) beyond the cut-off mark.

However, over the years quality of data under discussion in WDIs increased considerably and one may link the use of Gini coefficient obtained from microdata with direct examination of the different segments of the quintile income distributions. Figure 3 (as presented below) shows improved quality of data in WDI 2017, as the nature of the relationship and goodness of fit of the model go closely with those displayed in figure 1. In WDI 2017 data, shortfall ranges between 6.11 % (Kosovo) and 11.19 % (Seychelles).

[3] Including the two above-mentioned countries and excluding typos, such as in cases of Seychelles and Micronesia in WDI, 2010.

[4] http://data.worldbank.org/about/data-updates-errata (accessed on 30/03/2013).

Figure 3. Relationship between Gini coefficients obtained from microdata and quintile data as presented in World Development Indicators 2017; n= 161



4. Use of quintile data from UNU-WIDER World Income Inequality Database 2017

There are 5570 valid cases in WIID 3.4 showing quintile group-share of income or consumption. As above, we examine the relationship between the micro-Gini and those computed from quintile income distributions and present it in figure 4 below.

In figure 4, we see lots of distortion, some of which are due to purely misreporting and some are probably due to the inappropriate process of squeezing of microdata. Readers may take note of these cases while visual linking of the quintile distributions available in WIID 3.4 database with the respective micro-Ginis of the same database.

In this exercise, shortfall ranges from – 43.94 % to + 67.56 %, which is quite an absurd result. There are as many as 418 instances out of 5570 valid cases (7.5 %), which are unusual. We found negative shortfall for 89 cases out of 418 unusual cases. Negative shortfall means that after grouping of microdata, Gini coefficient increases. For example, in case of Ethiopia in 2010, Gini coefficient increases from 37.10 to 53.40 indicating a shortfall of - 43.94 %. Such a result is unexpected when the same microdata set is squeezed into quintiles. We may add another 161 cases (89+161 = 250), where the certain amount of underestimation (4 % in case of number of groups or k = 5) is less than 4 %. Also there are 168 cases, where shortfall exceeds 12 %. The highest shortfall arises for Zambia in 2004, the micro-Gini of which decreased from 55.00 to 17.84. Something went wrong with the cases of Zambia and others, which need to be addressed.

In order to check consistency of data according to source, we have formed two subsets of data respectively from WIID 3.4 with two popular sources: (i) the World Bank[5] (n = 1363), and (ii) the Luxembourg Income Study (n = 275). In case of the former, shortfall varies from 5.13 %

[5] Does not include 294 cases with source "Deininger & Squire, World Bank 2004", which may contain some discrepancies.

to 15.69 %, the final four cases of which seem unusual as shortfall exceeds 12 % [for example, Belize, 1993 (12.15 %); Namibia, 1993 (12.41 %); Belize, 1994 (12.45 %); and Malawi, 1997 (15.69 %)]. If we drop these four cases at this moment, we get a perfect straight-line relationship with an R-square value of 1.00 as shown above in figure 5.

Figure 4. Relationship between Gini coefficients obtained from microdata and quintile data as presented in WIID 3.4 (2017); n= 5536

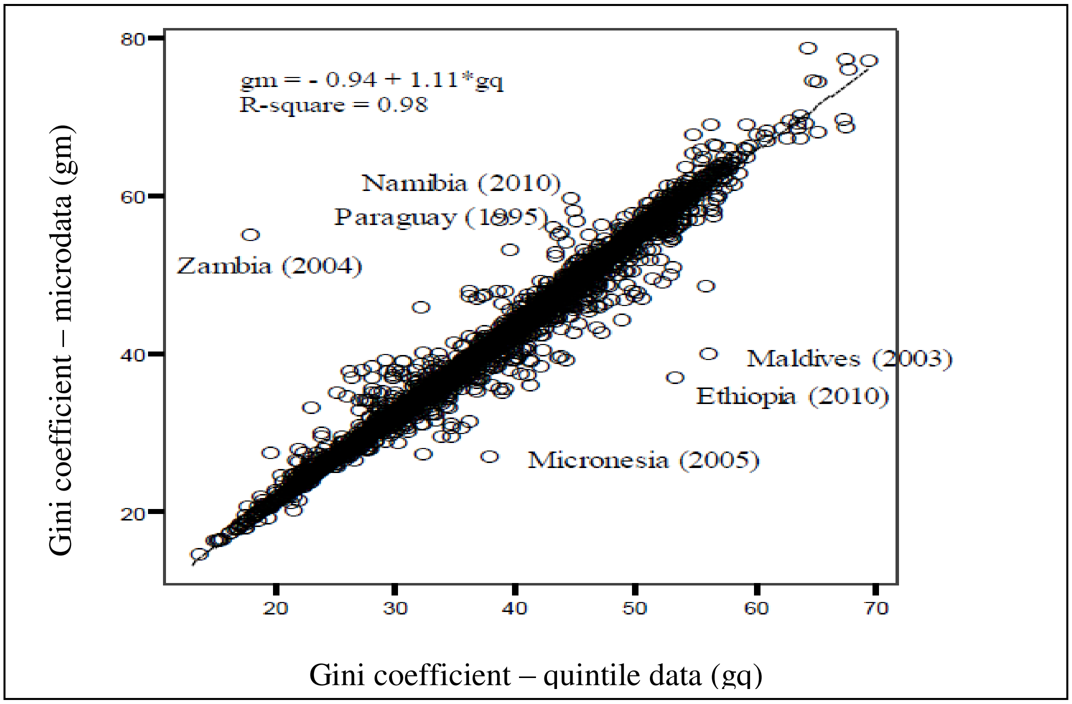
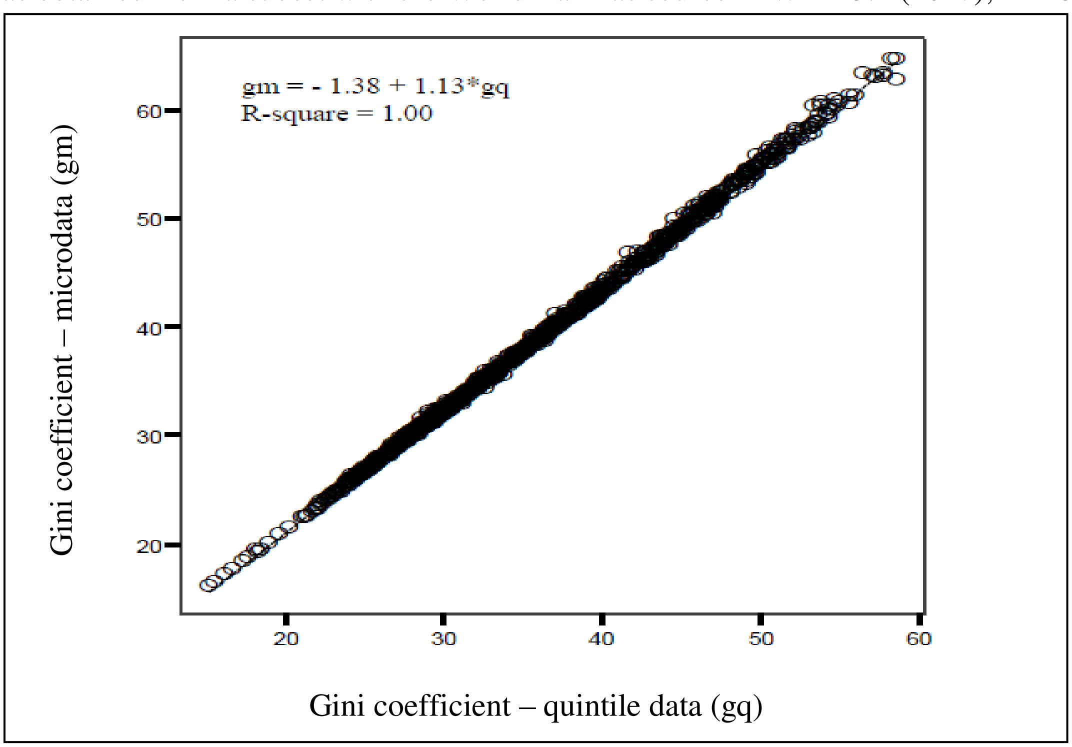
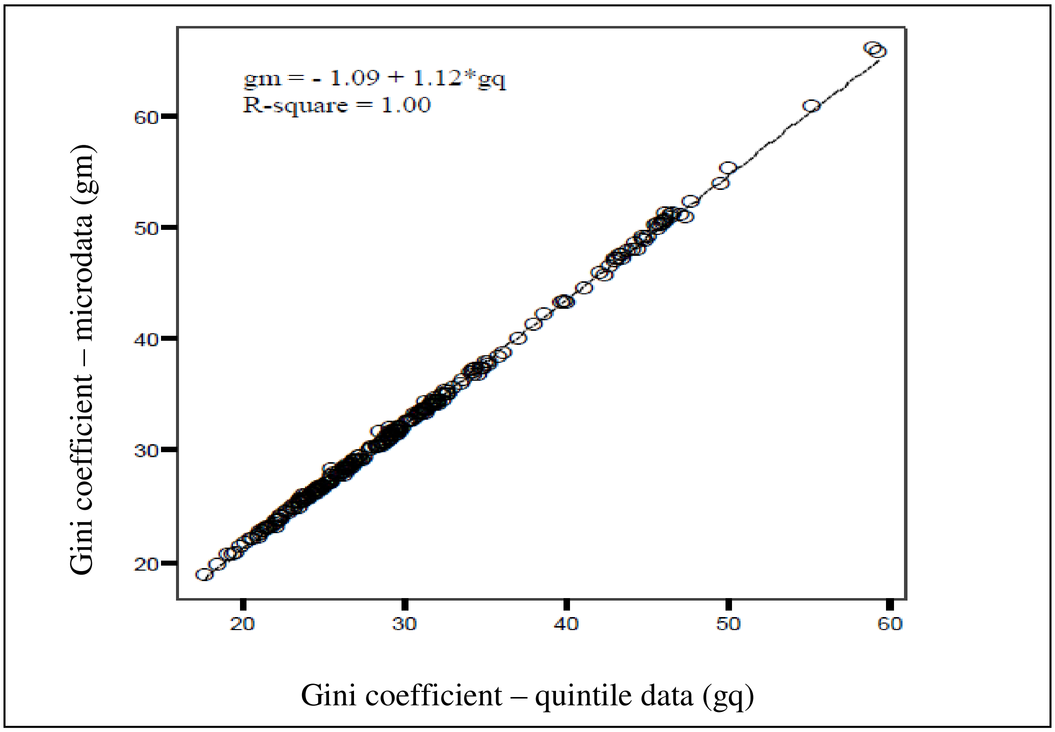


Figure 5. Relationship between Gini coefficients obtained from microdata and quintile data as obtained from a subset with the World Bank as source in WIID 3.4 (2017); n= 1359



The graph for the second subset of data with source as Luxembourg Income Study is also drawn as shown in figure 6 below. This relationship too is found perfect (and linear) with an R-square value of 1.00. In this subset of data, shortfall varies from 4.99 % to 10.76 %. So, in WIID 3.4, one may easily rely on data with source as the World Bank or the Luxembourg Income Study and use those to explain inequality conditions.

Figure 6. Relationship between Gini coefficients obtained from microdata and quintile data as obtained from a subset with the Luxembourg Income Study as source in WIID 3.4 (2017); n= 275



5. Use of decile data from UNU-WIDER World Income Inequality Database 2017

We repeat similar exercises for decile data in WIID 3.4. First, we compute decile Gini and then try to relate it with micro-Gini as reported in WIID 3.4. We form two subsets of data from WIID 3.4 (with 4958 valid cases) according to two major sources: (i) the World Bank[6] (n = 1308), and (ii) the Luxembourg Income Study (n = 276) and repeat the above-mentioned exercises.

We need to keep in mind, as reported by van Ourti and Clarke (2011), that a shortfall in case of grouping of microdata into deciles, may go up to 2.5 %, where the certain amount of underestimation is 1 %. In the present round of exercise, shortfall varies from - 49.26 % to + 65.71 %, which is again an absurd result. There are as many as 643 instances out of 4958 valid cases (≈ 13 %), which are unusual. There are 320 cases for each of which, shortfall becomes negative. In addition to these, there are 148 cases (320+148 = 468), where underestimation is below of 1 %. Also there are 175 cases, where shortfall exceeds 5 %.

In order to check consistency of data according to previously mentioned sources, we first present the results of data from the World Bank (in figure 8). We find a perfect (linear)

[6] Does not include 291 cases with source "Deininger & Squire, World Bank 2004", which may contain some discrepancies.

relationship between the micro-Gini and decile Gini with an R-square value of 1.00 (n = 1308). There are two distortions only (Malawi, 1997 and Namibia, 1993). We may ignore these two cases at this moment and derive conclusion from the rest of the result. We see that shortfall in this present round of exercise (for n = 1306, after ignoring the said two cases) ranges between 1.40 % and 5.00 %.

Figure 7. Relationship between Gini coefficients obtained from microdata and decile data as in WIID 3.4 (2017); n= 4958

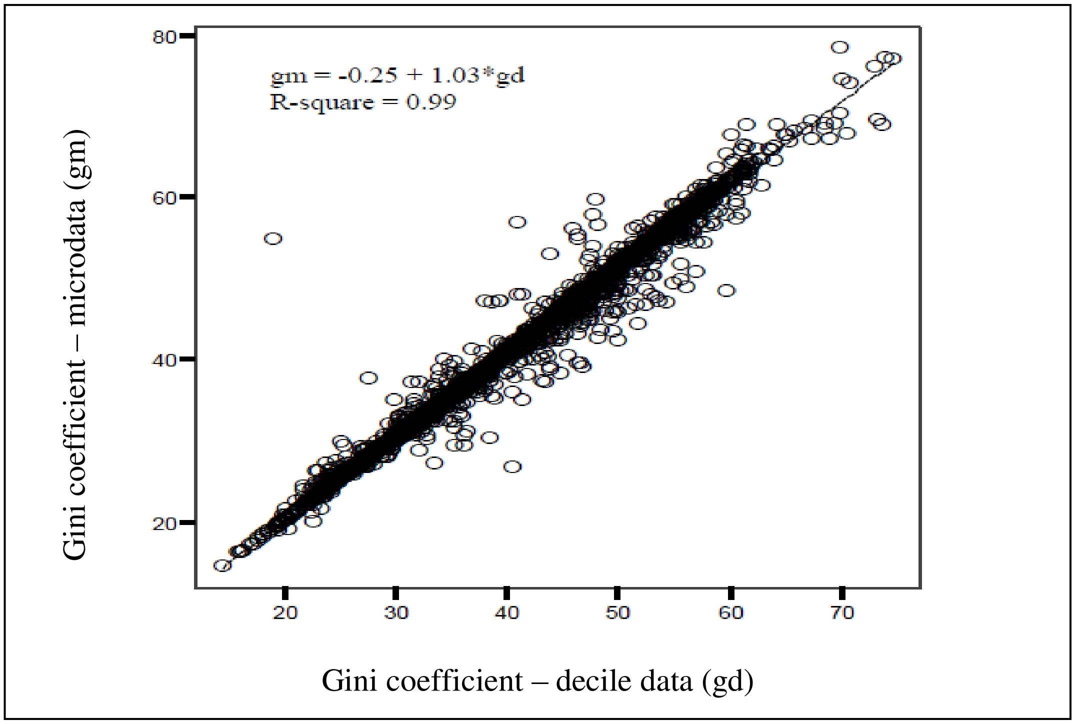
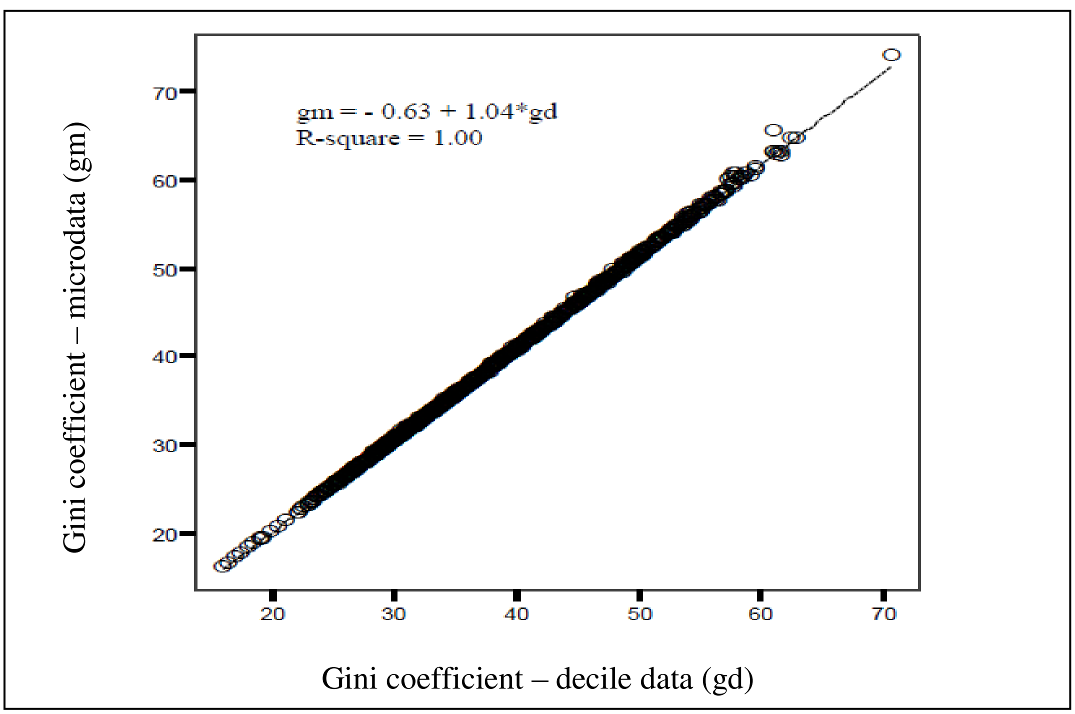
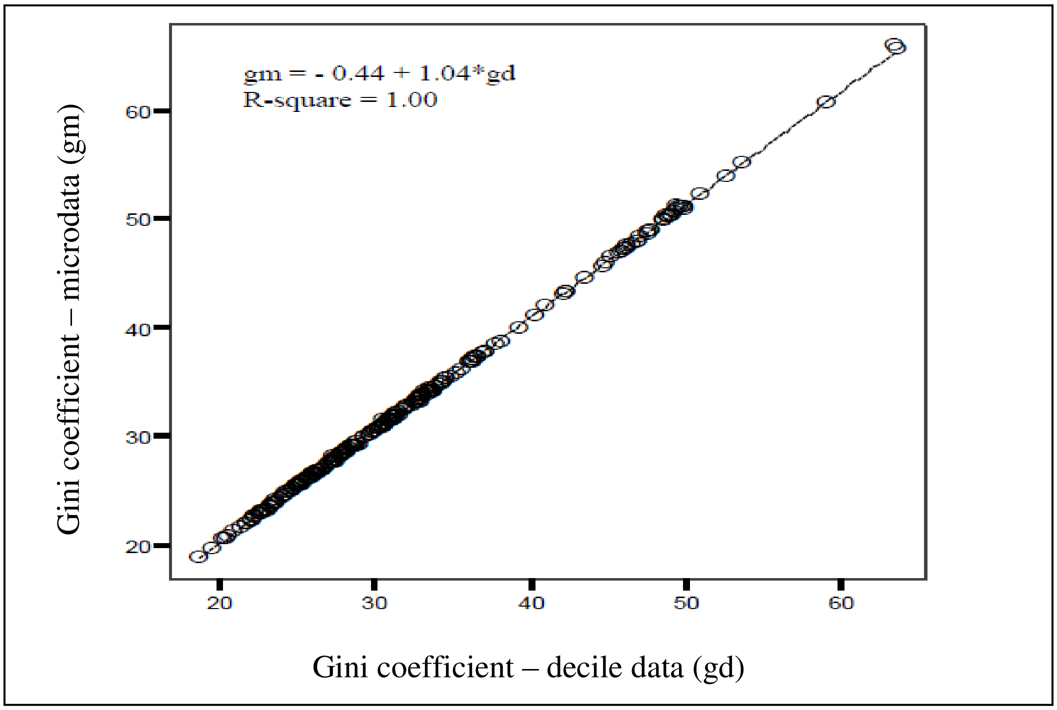


Figure 8. Relationship between Gini coefficients obtained from microdata and decile data as obtained from a subset with the World Bank as source in WIID 3.4 (2017); n= 1308



The relationship between micro-Gini and decile-Gini (with the Luxembourg Income Study as source of data) is far more accurate in figure 9 than in the previous one. Here shortfall varies from 1.17 % to 4.06 %. Combining the results obtained from these two immediate previous exercises, we may postulate that shortfall due to grouping for k = 10 may vary from 1 % to 5 %. So, in case of linking decile group-share of income with micro-Gini visually, one may keep in mind that if shortfall remains beyond this specified limit, there may exist discrepancy in data. In such a situation, one may avoid the cases under question in study.

Figure 9. Relationship between Gini coefficients obtained from microdata and decile data as obtained from a subset with the Luxembourg Income Study as source in WIID 3.4 (2017); n= 276



6. Summary of consistency check of data from some other sources in WIID 3.4

Besides the World Bank and the Luxembourg Income Study there are also some other official sources of data in WIID 3.4. We present summary of findings in regard to some of them briefly in table 4 below. Column 1 of the table shows some other sources of data in WIID 3.4. The second column shows number of cases in the data set. The third and fourth columns show the indicator (R-square value of linear relationship) of the first discussed method of checking consistency of microdata and squeezed data. The final two columns show the criteria of checking consistency of microdata and squeezed data under the second method. For example, the figure corresponding to ECLAC in the fifth column is 2; it implies that there are two cases for each of which the magnitude of shortfall is either less than 4 % or more than 12 %.

Table 4 is self-explanatory and shows that some particular cases [in regard to the (i) Socio-Economic Database for Latin America and the Caribbean (SEDLAC) 2016, (ii) European Commission, and (iii) Deininger & Squire, World Bank 2004] need attention. It tacitly implies that the displayed cases in the final two columns are to be dropped from study, at this moment, for correction or special attention.

Table 4. Summary of consistency check of data from some other sources in WIID 3.4

Sources	Number of cases	R-square value of linear relationship		Shortfall	
		Quintile data	Decile data	Number cases beyond the cut-off marks: 4 % to 12 %	Number cases beyond the cut-off marks: 1 % to 5 %
(1)	(2)	(3)	(4)	(5)	(6)
ECLAC	599	0.99	0.99	2	2
DIW	47	1.00	1.00	0	0
Socio-Economic Database for Latin America and the Caribbean (SEDLAC) 2016	714	0.98	0.96	17[a]	27[b]
Eurostat	365	0.99	0.99	0	0
European Commission	90	0.95	0.96	13[c]	41[d]
Deininger & Squire, World Bank 2004	292	0.94	0.94	121[e]	189[f]

[a] Contains 6 cases of positive bias; [b] Contains 14 cases of positive bias; [c] Contains 10 cases of positive bias; [d] Contains 26 cases of positive bias; [e] Contains 102 cases of positive bias; [f] Contains 187 cases of positive bias. All the cases of positive bias need theoretical attention.

Source: Self-elaboration of WIID 3.4 data

7. General reasoning and scope for further research

The effective application of the bias correction method suggested by van Ourti and Clarke (2011) to consistency checks of aggregated inequality data such as World Development Indicators (WDIs) and World Income Inequality Database (WIID) in addition to their empirical upper limits of admissible magnitude of shortfalls between quintile-/decile-Ginis and the corresponding micro-Ginis is believed to be helpful for researchers who conduct projects utilising those data.

Although our impression is that the method of van Ourti and Clarke (2011) is a seminal idea, it seemingly appears that their method is more suitable for the consistency checks suggested in the present paper rather than using it for estimation of Ginis from grouped income or consumption data originally suggested. This exercise probably put the method of van Ourti and Clarke (2011) on the most right track.

One possible cause for some inconsistencies (although we need further research) may be the existence of zero or negative income data. User Guide of WIID does not mention how zero or negative income data are processed by each data provider (except some French data for which population coverage is household with positive or zero taxable income). If considering income as a measure of standard-of-living, zero or negative values are not necessarily preferable. Such values may be excluded when calculating micro-Ginis while those may be included when compiling quintile and decile figures. If so, micro-Ginis may possibly be lower than quintile-/decile-Ginis. However, we keep this issue open for further research.

8. Conclusion

As the thrust on alternative ways to explain inequality conditions is tending to increase either by supplementing or by replacing the use of Gini coefficient by visual examination of the income distributions, consistent data should be made readily available in the popular databases for the use of the common readers. Number of observations in each case in microdata may also be reported. In order to develop simple methods for consistency check of squeezed data, we discussed about issues like shortfall, underestimation, (downward) bias etc. As observed from the preliminary results, there are nearly 7.5 % and 13 % cases, which are unusual corresponding to quintile and decile data respectively in WIID 3.4. Some of these are simply misreporting or typos, which are to be corrected, and some appeared to be very special (with positive bias instead of downward bias), which warrant theoretical attention for further research. Common readers may restrain themselves from using the unusual cases in study till we resonate to further clarification.

References

Lerman, R. and S. Yitzhaki. (1989) "Improving the Accuracy of Estimates of Gini Coefficients" *Journal of Econometrics* **42**, 43-47.

Majumder, A. (2010) "A note on Gini index of Liberia in World Development Indicators 2009" available at: https://amlan.co.in/ alternative-measures-of-inequality/ (accessed on 06/09/2017).

Milanovic, B. (2012) "Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates" *Journal of Economic Inequality* **10**, 1-18.

Osberg, L. (1981) *Economic Inequality in Canada*. Butterworths: Toronto.

Osberg, L. (2017) "On the Limitations of Some Current Usages of the Gini Index" *Review of Income and Wealth* **63**, 574-584.

Piketty, T. (2014) *Capital in the Twenty-First Century*, Harvard University Press: Cambridge.

van Ourti, T. and P. Clarke. (2011) "A simple correction to remove the bias of the Gini coefficient due to grouping" *The Review of Economics and Statistics* **93**, 982-994.

UNU-WIDER. (2017) *World Income Inequality Database (WIID3.4), January 2017*. UNU-WIDER: Luxembourg.