WORKING PAPER | ISSUE 8 | 2016

Some are more equal than others: new estimates of global and regional inequality

ZSOLT DARVAS

ABSTRACT

We compare four methodologies to estimate the global distribution of income and find that many methods work well, but the method based on two-parameter distributions is more accurate than other methods. This method is simpler, easier to implement and relies on a more internationally-comparable dataset of national income distributions than other approaches used in the literature to calculate the global distribution of income. We suggest a simulation-based technique to estimate the standard error of the global Gini coefficient. Global income inequality among the citizens of 128 countries gradually declined in 1989-2013, largely due to convergence of income per capita, which was offset by a small degree the increase in within-country inequalities. The standard error of the global Gini coefficient is very small. After 1994, market income inequality in the EU28 was at a level similar to market inequality in other parts of the world, but net inequality (after taxes and transfers) is at a much lower level and it declined between 1994 and 2008, since when it remained relatively stable. Regional income inequality is much higher in Asia, Africa, the Commonwealth of Independent states and Latin America than in the EU28. In Asia, regional inequality has increased recent years, while it declined in the other three non-European regions.

Zsolt Darvas (zsolt.darvas@bruegel.org) is Senior Fellow at Bruegel and a Senior Research Fellow at the Research Centre for Economic and Regional Studies of the Hungarian Academy of Sciences and at Corvinus University, Budapest.

This paper is a background paper to Zsolt Darvas and Guntram B. Wolff (2016) 'An anatomy of inclusive growth in Europe,' Bruegel Blueprint 26, supported by the MasterCard Center for Inclusive Growth. The author is grateful to Bruegel colleagues and participants in the 2016 meeting of Project LINK for comments and suggestions, and to Uuriintuya Batsaikhan and Jaume Martí Romero for research assistance.

bruegel

1. Table of Contents

2. Introduction

Indicators of income distribution, such as quantile income shares and the Gini coefficient, are available for individual countries, but from official statistical sources they are not available for the world as a whole or for various country groups, such as the European Union (EU). While Eurostat publishes Gini coefficients for 28 EU countries and for various groups of countries within the EU, these Gini coefficients are population-weighted averages of country-specific Gini coefficients. However, the average of the Gini coefficients of individual countries does not correspond to the Gini coefficient of the combined population of those countries, partly because of the differences in average income in different countries, and partly because of differences in within-country income distributions[1].

The straightforward way to calculate the global distribution of income would be to pool together income data from all households in all countries to obtain the income distribution of all the world's households. This pooled distribution could be used to calculate the Gini coefficient and other indicators of income inequality. Unfortunately, such household-level income data is not available.

A number of academic works have estimated the global distribution of income. These works approximate more detailed data points on the country-specific income distributions (eg the 100 percentiles) than what is published by statistical offices (eg the five quintiles). Then, using a measure of average income and population size, they combine the detailed country-specific income distributions into a global distribution of income.

Two major data types were used in the literature for the estimation of more detailed information on country-specific income distributions.

Several authors, such as Bourguignon and Morrisson (2002), Bhalla (2002), Milanovic (2002), Morrisson and Murtin (2004) and Sala-i-Martin (2006), use quantile data from household surveys, such as deciles, quintiles or whatever quantile information is available. One of the biggest problems with such an approach is the lack of comparability between national surveys. Subsequently, the missing data has to be approximated, which can present other significant problems. In Europe, Eurostat quantile data, which allows for cross-country comparisons, is available for only a rather short period for all (or most) EU countries. Data for all current 28 EU members is available only from 2010, while data for all the first 15 EU members is only available from 2005. One may look to other sources for earlier data, but availability and comparability of such data is not ensured, to say nothing of the time-consuming process it requires to obtain this data.

In contrast, Chotikapanich, Valenzuela and Rao (1997) assume that within-country distributions follow the log-normal distribution (with different parameters in different countries) and use only

[1] A simple example illustrates the importance of differences in average income across countries. Suppose there is a country in which everyone earns the same and therefore there is no inequality (the Gini coefficient is zero). Suppose there is another country in which there is also no inequality. There is inequality if the two countries are considered jointly if the average income is different in the two countries and thereby the Gini coefficient (non-zero) for the two countries together is not the average of the Gini coefficients of the two countries (which are both zero).

the country-specific Gini coefficient and mean income to estimate the parameters of this distribution. Therefore, a key advantage of this method is that it does not require detailed data on income distribution, but only the Gini coefficient. A possible problem with this approach is that log-normal distributions might not describe the distribution of income in all countries very well.

In this paper we analyse the accuracy of various methods in the particular cases of four countries: the United States, Australia, Canada and Turkey. The national statistical offices of all four countries make both territorial (ie state-level) and country-wide income distribution data available. Thus, using data from the 50 US states and Washington DC, the 8 Australian states and territories, 10 Canadian provinces and 12 Turkish regions, we can calculate exactly how accurate the various methods are in estimating the country-wide Gini coefficient. We also assess the accuracy of various methods using quantile data from Eurostat for European countries. We find that many methods work quite well if the right level of detail is used about quantile income shares. In the end, however, we find that methods based on two-parameter distributions are among the most accurate.

We develop this method further using a stochastic simulation technique, which allows the calculation of a confidence band for the global Gini coefficient. In essence, our method involves simulating artificial samples of household income in each country so that the expected value of the Gini coefficient equals the Gini coefficient observed in the actual data and the expected value of the mean income equals the mean income observed in the actual data. We rely on the easily accessible and internationally comparable data on country-specific Gini coefficients from the Standardised World Income Inequality Database (SWIID) of Solt (2016). This dataset includes information on the uncertainty of (country-specific) Gini coefficients that we use to estimate the uncertainty of the global Gini coefficient. For the simulations we use random numbers generated from statistical distributions which were found to describe income distributions well: the log-normal distribution, the Pareto distribution and the Weibull distribution. Once artificial samples of household incomes are simulated for each country, we then pool these simulated household incomes data for all countries into a single sample to obtain the income distribution of global citizens and calculate the global Gini coefficient and other indicators of inequality and poverty.

Section 2 reviews existing methodologies for calculating the Gini coefficient for world citizens, followed by our proposal to extend the two-parameter based method in section 3. Section 4 compares the ability of various methods to estimate the overall US, Australian, Canadian and Turkish Gini coefficients from territorial (ie state-level) data of these countries, analyses the robustness of the methods based on quantile incomes shares to the level of data detail, and compares the similarity of the estimates by various methods. Section 5 presents our global and regional Gini coefficient estimates for 128 countries and five main regions (Asia, Africa, Commonwealth of Independent States, the EU and Latin America) for the 1989-2013 period for the world and most regions, and for 1989-2015 for the EU. This section also decomposes the change in the global and regional Gini coefficients to within-country inequality changes and other factors. Section 6 concludes.

Our global and regional Gini coefficient estimates are downloadable from: http://bruegel.org/publications/datasets/global-and-regional-gini-coefficients/. We plan to update our estimates when updated data on country-specific Gini coefficients becomes available.

3. Earlier methods for estimating the world distribution of income

A number of attempts have been made to approximate the world distribution of income and to calculate statistics of global income inequality. Since household-level data is not available worldwide and national statistical offices publish only a few aggregate indicators of within-country inequality, the first challenge is how to approximate more detailed data on income distribution within each country beyond what is available.

Chotikapanich, Valenzuela and Rao (1997) highlighted some of the problems with survey-based data. They argued that the log-normal distribution describes within-country income distributions accurately and recognised that the two parameters of this distribution can be identified with the Gini coefficient and mean income. They estimate the parameters of the log-normal distribution for each country.

Many other papers use quantile data on income shares:

- Identical quantile income method: Bourguignon and Morrisson (2002) and Milanovic (2002) assume that each quantile in a country is made up of individuals with identical incomes[2]. For example, all people belonging to the bottom 10 percent of the income distribution in a given country are assumed to have the same income. Countries differ in terms of the available detail on quantile income shares, eg for some countries only quintile shares are available, while for others data on deciles, or even more detailed information is available. Ideally, this methodology should use the most detailed quantile data.
- Lorenz-curve regression method: Bhalla (2002), building on Kakwani (1980), adopts a regression method to approximate the Lorenz-curve in each country based on the limited number of quantile income share data available[3]. The estimated regression proposed by Kakwani (1980) is the following:
  $$\log[p - L(p)] = \beta_1 + \beta_2 \log p + \beta_3 \log(1 - p),$$
  In which p represents the bottom p percent of the population, $L(p)$ is the corresponding share in income (ie the value of the Lorenz-curve at p), while β_1, β_2 and β_3 are parameters to be estimated. Bhalla (2002) then uses the estimated regression to project the Lorenz-curve at the 100 percentiles of the income distribution for each country, plus makes some adjustments to ensure that the final set of the 100 percentiles used are consistent with available data on income shares (eg the sum of the first 20 percentiles is the same as the data on income share of the lowest quintile, etc).
- Kernel density method: Sala-i-Martin (2006) first assumes that individuals belonging to each quintile have identical incomes, which allows him to draw the histogram of incomes as five equal-height bars at the estimated mean income of people belonging to each of the five quintiles. After taking logs, he then uses a non-parametric kernel

[2] Milanovic (2002) acknowledges that the same method has been used by several previous works during the preceding two decades.
[3] Bhalla (2002) calls this regression method the 'Simple Accounting Procedure' (SAP), yet we find the name 'Lorenz-curve regression method' more accurate.

function to estimate the 100 percentiles of the empirical density function of each country's income distribution.

- Beta distribution: Chotikapanich, Griffiths, Rao and Valencia (2012) estimate the three parameters of the beta distribution (for each country) using a method-of-moments estimator based on data of income shares.

Once the 100 percentiles of the income distribution are estimated, a measure of mean income is used to estimate the incomes of households corresponding to the 100 percentiles of the income distribution. Two main measures of mean income were used:

- GDP per capita at purchasing power parity (PPP) (eg Chotikapanich, Valenzuela and Rao, 1997; Bourguignon and Morrisson, 2002; Bhalla, 2002; and Sala-i-Martin, 2006; Chotikapanich, Griffiths, Rao and Valencia, 2012);
- Mean income or mean expenditure from surveys converted to a common numeraire by using PPP exchange rates (eg Milanovic, 2002).

The advantages of GDP per capita are its comparability across countries and its availability for a wide range of countries and historical periods. However, GDP per capita is an imperfect proxy of mean household income, because of the inclusion of non-household incomes in GDP. In principle, data on mean household income should be used. Unfortunately, it is not available for all countries, since in the surveys of several countries only mean expenditures (and not mean incomes) are available. The definition of income and expenditure also varies in different countries. Chotikapanich, Griffiths, Rao and Valencia (2012) collected data both on GDP per capita and on mean incomes/expenditures and decided to use GDP per capita. Their main arguments for this choice were (a) comparability problems with mean income and expenditure data across countries, (b) GDP per capita is a widely-used broad measure of standard of living, and (c) GDP per capita is easily available for a large number of countries.

Finally, by using the population size of each country, the approximated incomes of individuals in each country are pooled together to get the world distribution of income[4]. This world income distribution is then used to calculate various indicators of inequality, including the Gini coefficient.

The above-mentioned six works all estimate the Gini coefficient in 1970-2000 to be near 65, with a small decline in the 1990s (Table 1), despite the differences in approximating within-country income distributions and mean incomes and differences in the composition and number of countries considered[5]. Most likely, global inequality is primarily driven by between-country inequality, and thus within-country inequality (and the way within-country income distribution is approximated) is less relevant. We test this hypothesis in section 5.

---

[4] Some of the papers adopt slightly different steps to calculate the world distribution of income, yet the essence of all approaches is the same.

[5] The results of these studies are broadly comparable, because they are based on data that was available around 2000. Since then, major revisions to purchasing power exchange rates have occurred, which alter the results. Chotikapanich, Griffiths, Rao and Valencia (2012) note that the use of the new PPP exchange rates increases the estimated global Gini coefficient by about several points.

Table 1: Some earlier estimates of the global Gini coefficient

Authors	Method for within-country income distribution	Income distribution data	Income measure	Global Gini coefficient						
				1970	1980	1988	1990	1992	1993	2000
Chotikapanich, Valenzuela and Rao (1997)	Log-normal distribution	Gini	GDP per capita		65.8		64.8			
Chotikapanich, Griffiths, Rao and Valencia (2012)	Beta distribution based	Income shares	GDP per capita						64.8	64.0
Bhalla (2002)	Lorenz-curve regression method	Income shares	GDP per capita	≈ 68.7	≈ 68.6	≈ 67.2	≈ 67.5	≈ 67.2	≈ 67.0	≈ 65.2
Bourguignon and Morrisson (2002)	Identical quantile income method	Income shares	GDP per capita	65.0	65.7			65.7		
Milanovic (2002)	Identical quantile income method	Income shares	Income from surveys			62.5			65.9	
Sala-i-Martin (2006)	Kernel density method	Income shares	GDP per capita	65.3	66.0	64.9	65.2	64.5	64.0	63.7

Sources: Table 1 of Chotikapanich, Valenzuela and Rao (1997), Table 8 of Chotikapanich, Griffiths, Rao and Valencia (2012), Figure 11.1 of Bhalla (2002), Table 1 of Bourguignon and Morrisson (2002), Table 16 of Milanovic (2002) and Table III of Sala-i-Martin (2006). Note: the country coverage in each of these works was different.

4. Extending the method based on two-parameter distributions

Chotikapanich, Valenzuela and Rao (1997) use the two-parameter log-normal distribution to approximate within-country income distribution in a deterministic setting. We extend this method by considering other distributions and a stochastic setting too.

Various articles have found that income distribution within a country can be well approximated by a number of parametric statistical distributions. Nice summaries of this literature are presented in Cowell (2009) and Lubrano (2015). These authors conclude that two-parameter distributions, and their mixtures, are the most useful for modelling incomes, while they are sceptical about the use of more complicated distributions with three or four parameters. Thus we use three two-parameter distributions: the log-normal distribution, the Pareto distribution and the Weibull distribution. Two-parameter distributions are especially appropriate for our study, given that we wish to use two indicators (mean income and the Gini coefficient) to set the parameters of the distribution. The probability density function, mean and the Gini coefficient derived from these distributions are included in Table 2.

Table 2: Probability density function, mean and the derived Gini coefficient of three distributions we use

	Probability density function	Mean	Gini coefficient
Log-normal	$f(x)=\frac{1}{x\sqrt{2\pi s^2}}e^{-\frac{(\ln x-m)^2}{2s^2}}, x>0$	$\mu=e^{m+\frac{s^2}{2}}$	$G=2\Phi\left(\frac{s}{\sqrt{2}}\right)-1$
Pareto	$f(x)=\frac{ab^a}{x^{a+1}}, x>b, a>0$	$\mu=\frac{ab}{a-1}, a>1$	$G=\frac{1}{2a-1}, a>1/2$
Weibull	$f(x)=hk^{-h}x^{h-1}e^{-\left(\frac{x}{k}\right)^h}, h,k>0, 0<x<\infty$	$\mu=k\Gamma\left(1+h^{-1}\right)$	$G=1-2^{-1/h}$

Source: Lubrano (2015) and http://mathworld.wolfram.com/.

Note: $\Phi(.)$ in expression for the Gini coefficient of the log-normal distribution is the cumulative distribution function of the standard normal distribution. $\Gamma(.)$ in the expression for the mean of the Weibull distribution is the gamma function.

Data on the Gini coefficient allows the calculation of one parameter of the distribution (*s* for log-normal, *a* for Pareto and *h* for Weibull), while this parameter and data on mean income allows a calculation of the second parameter of the distribution (*m* for log-normal, *b* for Pareto and *k* for Weibull), for each country and for each year.

After obtaining the parameters, these distributions can be used to describe within-country income distribution. In a deterministic setting, the cumulative distribution function (in conjunction with population size) can be used to approximate individual incomes.

A stochastic approach based on random number generators can also be useful, for two reasons. First, these distributions may not describe income distributions perfectly, in which case any random sample from these distributions would be equally likely. Second, we wish to estimate the standard error of the global Gini coefficient. Our data source for the Gini coefficient, the Standardised World Income Inequality Database (SWIID) of Solt (2016), includes information about the uncertainty of the (country-specific) Gini coefficients. We can incorporate this uncertainty into the calculation of the global Gini coefficient.

Our stochastic approach is based on random number generators from the parametric distributions. We use random numbers to simulate artificial samples of household income in each country so that:

- The expected value of the Gini coefficient equals the Gini coefficient observed in the actual data in each country, and
- The expected value of the mean income equals the mean income observed in the actual data in each country.

For each country and year, we simulate artificial household income data proportional to the population. For example, for Germany, the EU country with the largest population of about 82 million in 2010, we simulate about 82,000 artificial income data points in 2010. For Malta, the EU country with the smallest population, we simulate about 400. We then pool the simulated household income data from all countries into a single sample to approximate the global (or regional) distribution of income. For example, for the EU, we simulate approximately 501,000 data points (corresponding to the 501 million inhabitants in the 28 EU countries) for 2010. We then calculate the Gini coefficient from this set of combined income distributions of households of the countries considered.

We use two versions of the stochastic method, depending on whether or not information about the uncertainty of the Gini coefficient is incorporated:

- Simple version: we just use the published Gini coefficient (or the mean of the 100 iterations included in the SWIID) to calibrate the parameters of the distribution.
- Full version: we incorporate the uncertainty in country-specific Gini coefficients using the SWIID. This dataset includes 100 iterations for the Gini coefficient of each country, reflecting the uncertainty in the Gini coefficient estimate. According to Solt (2016), the 100 iterations for the different countries are independent from each other. Therefore, we sample without replacement from the 100 iterations for each country to obtain a particular realisation of the Gini coefficient. For different countries, we draw from the 100 country-specific iterations independently from each other. For example, we may draw the 6th iteration for country A, the 87th for country B, the 55th for country C, and so on. For a particular drawing of country-specific Gini coefficients, we calculate the corresponding global Gini coefficient using a two-parameter distribution method. Next, we draw again a new set of country-specific Gini coefficients and calculate again the corresponding global Gini. And so on: we do altogether 100 drawings and thereby we use all country-specific Gini coefficient iterations included in the SWIID database but most likely in a different order across countries. This procedure can capture the uncertainty of the global Gini coefficient related to the country-specific Gini coefficients, yet we cannot incorporate the uncertainty related to the mean income of the countries. After obtaining 100 estimates for the global Gini coefficient, we report the mean and the standard deviation across the 100 estimates. The 100 estimates are available in the dataset that can be downloaded from Bruegel's website.

The method based on two-parameter distributions is simple, easy to implement, and is based on an easily accessible and internationally comparable dataset of (country-specific) Gini coefficients. To our knowledge, the Standardised World Income Inequality Database is the most comprehensive dataset of Gini coefficients aimed at maximising comparability and providing the broadest possible coverage across countries and years. The use of this dataset also allows rather long sample periods to be studied. For example, we calculate global and regional Gini coefficients for the 1989-2013 period[6]. In contrast, Eurostat data on quantile income shares of the current 28 member of the European Union is available only starting in 2010, for 27 countries (not including Croatia) from 2007, and for 25 countries (not including Croatia, Romania and Bulgaria) from 2005. Therefore, consistent data on quantile income shares, which is needed for the other the methods reviewed in the previous section, is available from Eurostat for a much shorter period.

[6] For the EU, we use the 1989-2015 period.

5. Testing the methodologies

4.1 The perfect aggregation test: estimating the US, Australian, Canadian and Turkish Gini coefficients from territorial data

There is a perfect test for the accuracy of the various methodologies in the particular cases of those countries for which data on income distribution (quantile income shares and Gini coefficient), mean income, and population are available at territorial level as well as for the country as a whole. Thereby, we can perfectly check the accuracy of the methodologies in estimating the country-wide Gini coefficient from territorial data and compare the estimates to the country-wide data published by the statistical offices. The estimation of the global and European Gini coefficients from country data is done in exactly the same way as the estimation of the country-wide Gini coefficient from the territorial data of the four countries.

We therefore collected territorial (sub-federal and regional) and country-wide data for four countries: United States (50 US states and DC), Australia (8 states and territories), Canada (10 provinces[7]) and Turkey (12 regions).

The following quantile income shares are available at the territorial level (as well as at the country level) for the four countries (see data sources in the Annex):

- USA: quintile income shares and the top 5% income share;
- Australia: quintile income shares;
- Canada: decile income shares;
- Turkey: decile income shares.

For better comparability of the results for the four countries, we report results that are based on quintile income shares only for all four countries. For the US, Canada and Turkey, we also report results using the additional quantile shares data available.

Figure 1, Figure 2, Figure 3 and Figure 4 show, based on territorial data, the estimated country-wide Gini coefficients derived from the various methods in each year, as well as the actual country-wide data as published by the statistical offices of these countries. Table 3, Table 4, Table 5 and Table 6 summarise the results by presenting the average absolute deviation of the estimates from the known country-wide data through the years. A number of interesting conclusions can be drawn out.

First, both the weighted and the unweighted average of territorial Gini coefficients are well below the actual data for the country as a whole for all four countries. This finding suggests that the Eurostat Gini coefficient data for EU and euro-area aggregates, which are population weighted averages of country-specific Gini coefficients, are likely to underestimate the true Gini coefficient for EU and euro-area citizens.

[7] Canada consists of 10 provinces and three territories. Income distribution data is not available for the three territories, but since these three territories account for only about 1.0-1.5 percent of total Canadian population, their omission in our calculation is a minor issue.

Second, several methods are surprisingly good at estimating the country-wide Gini coefficient from territorial data. As Table 3 indicates for the US, the average absolute error of the best methods in 2006-2014 is a mere 0.03, very small compared to the typical Gini values of 47 in the US. The estimation errors of the best methods are also quite small at about 0.1 in Canada and 0.3 in Australia, against their near-average Gini coefficients around 30, and also about 0.1 in Turkey, where the Gini coefficient is about 40.

Third, methods based on two-parameter distribution appear to work very well. These methods are among the most accurate methods. Even the best method for the US, Australia and Canada is based on a two-parameter distribution, while for Turkey it is the second best. It does not seem to matter much whether we use the log-normal, the Pareto or the Weibull distribution. In the cases of the US and Australia, however, the deterministic method-based Pareto distribution has led to somewhat higher estimation errors, although this is the most accurate method for Canada and Turkey. It also does not seem to matter much whether we use a deterministic or stochastic approach at least for the log-normal and Weibull distributions, while for the Pareto distribution there were some differences[8].

Fourth, among the methods using quantile data, the Lorenz-curve regression methods of Kakwani (1980) and Bhalla (2002) seems to be the most robust[9]. In the cases of all four countries this method is rather precise irrespective of whether only quintile income shares or more detailed income shares data are used. In contrast, the identical quantile income method of Bourguignon and Morrisson (2002) and Milanovic (2002) works poorly for all countries when only quintile income shares are used: it severely underestimates the country-wide Gini coefficient. This method works much better when data on the top 5 percent income share is also used for the US and the top 10 percent income share for Canada and Turkey, underlining that the distribution within the top 20 percent has a major impact on the Gini coefficient. The Kernel density method of Sala-i-Martin (2006) works quite well when only quintile data is used (as in Sala-i-Martin, 2006), but this method performs much worse when additional quantile information is added[10]. It may sound puzzling that a method produces worse results when more detailed data is used. Since the Kernel function smooths out income shares both up and down, when information on top 5 percent (US) or top 10 percent (Canada and Turkey) income shares is added, this method may smooth upward too much.

Certainly, while our calculations for the US, Australia, Canada and Turkey are reassuring, they do not prove that these methods work well for other countries or for groups of countries.

---

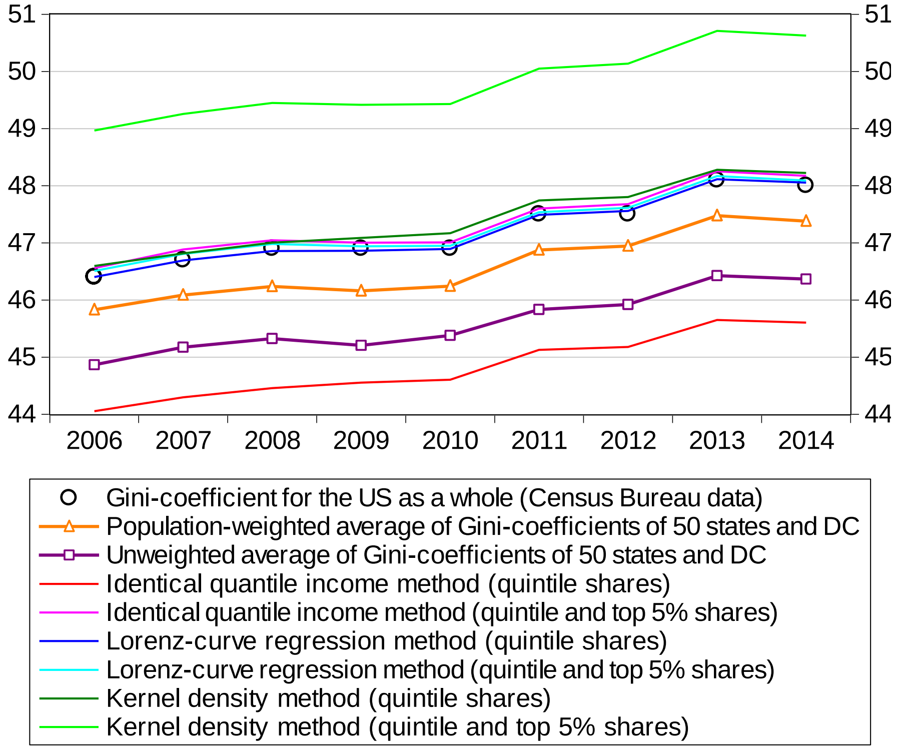
[8] For the stochastic method, we use the simple version described in the previous section due to data availability issues.

[9] As we noted in Section 2, after estimating the regressions, Bhalla (2002) made some adjustments to ensure that the final set of the 100 percentiles used is consistent with available data on income shares. We did not incorporate these adjustments, because the method without the adjustment already works well. Thus, we essentially used the method of Kakwani (1980).

[10] Like Sala-i-Martin (2006), we estimate the Kernel-function on logarithmic income. Interestingly, the method is less accurate when the Kernel function is estimated on actual (not log) data. Sala-i-Martin (2006) used the same bandwidth for all countries and years, which he calibrated on the basis of the standard formula: $w = 0.9 * \sigma * n^{-0.2}$, where w is the bandwidth for the Kernel, σ is the standard deviation of log-income and n is the number of observations. He calibrated the bandwidth by assuming an average value for the standard deviation. Instead, we select the bandwidth for each country and year with the standard formula, because there were major differences in the standard deviation of log-incomes across the countries.

Figure 1: The overall US Gini coefficient and its estimates from data of 50 states and DC, 2006-2014

(A) Methods based on income share data



(B) Methods based on two-parameter distributions

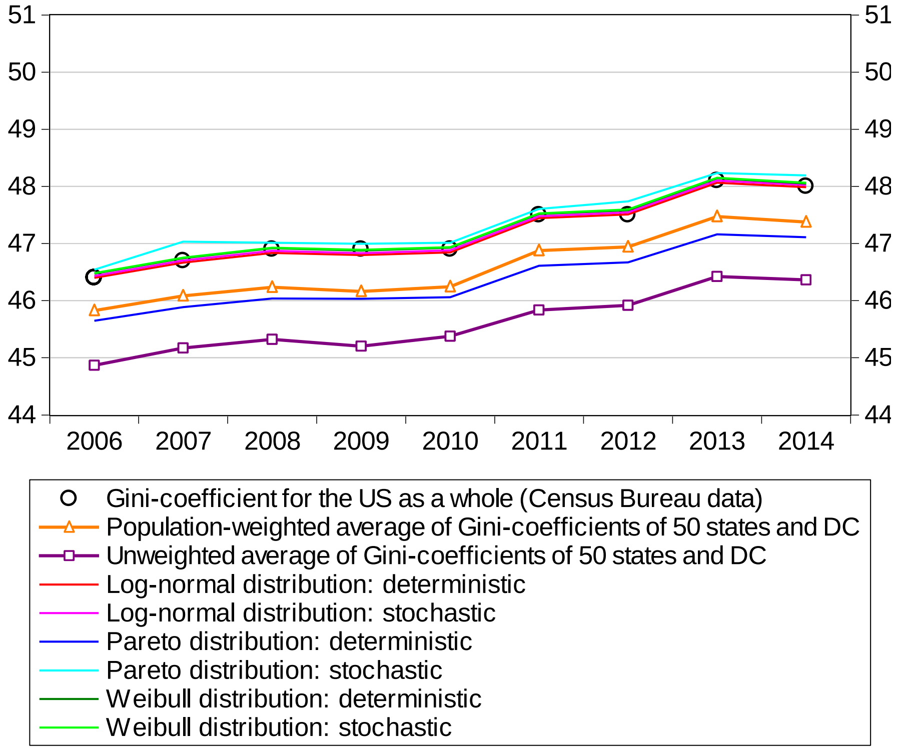
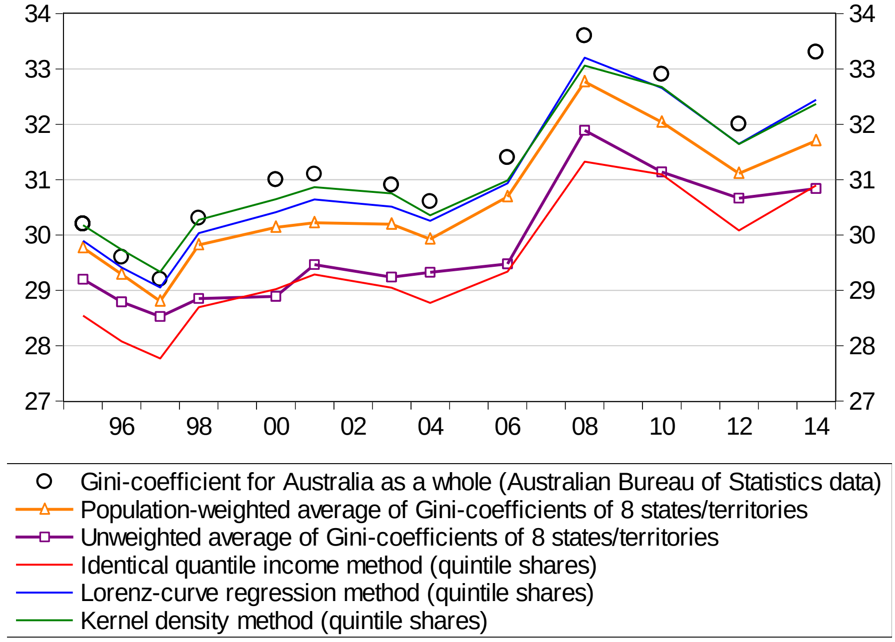
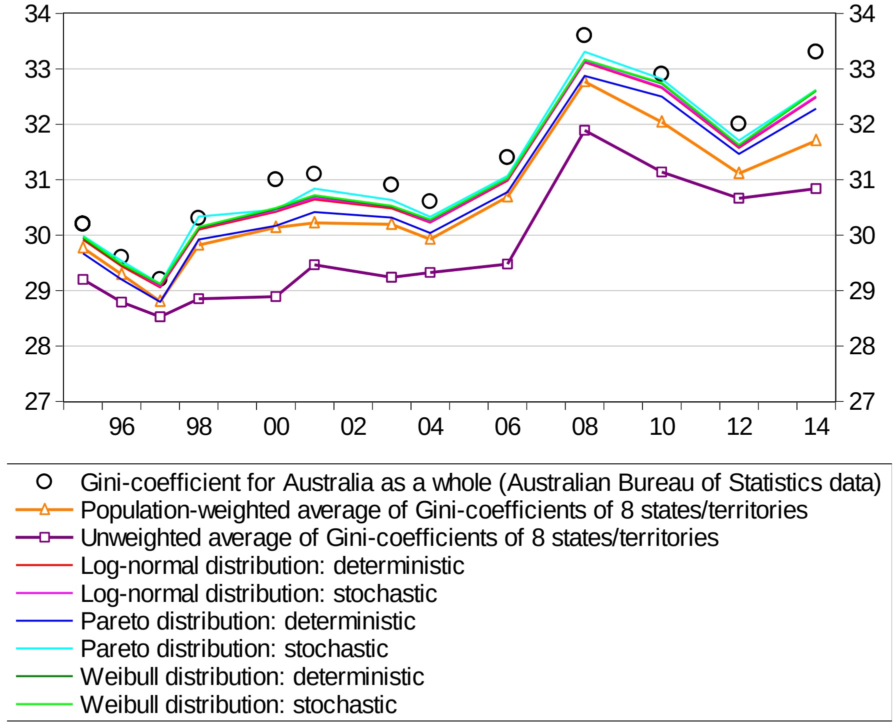


Figure 2: The overall Australian Gini coefficient and its estimates from data of 8 states and territories, 1995-2014

(A) Methods based on income share data



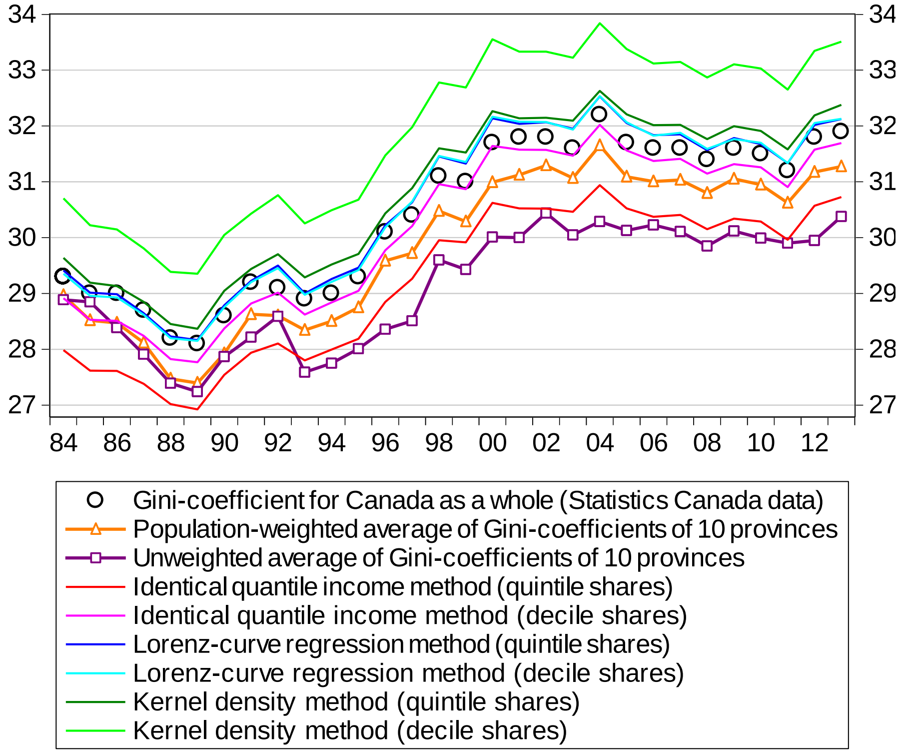
(B) Methods based on two-parameter distributions



Note: several surveys were conducted in 2-year periods that we report at the second years. We connect all lines (except for the actual Gini coefficient) for better readability.

Figure 3: The overall Canadian Gini coefficient and its estimates from data of 10 provinces, 1984-2013

(A) Methods based on income share data



(B) Methods based on two-parameter distributions

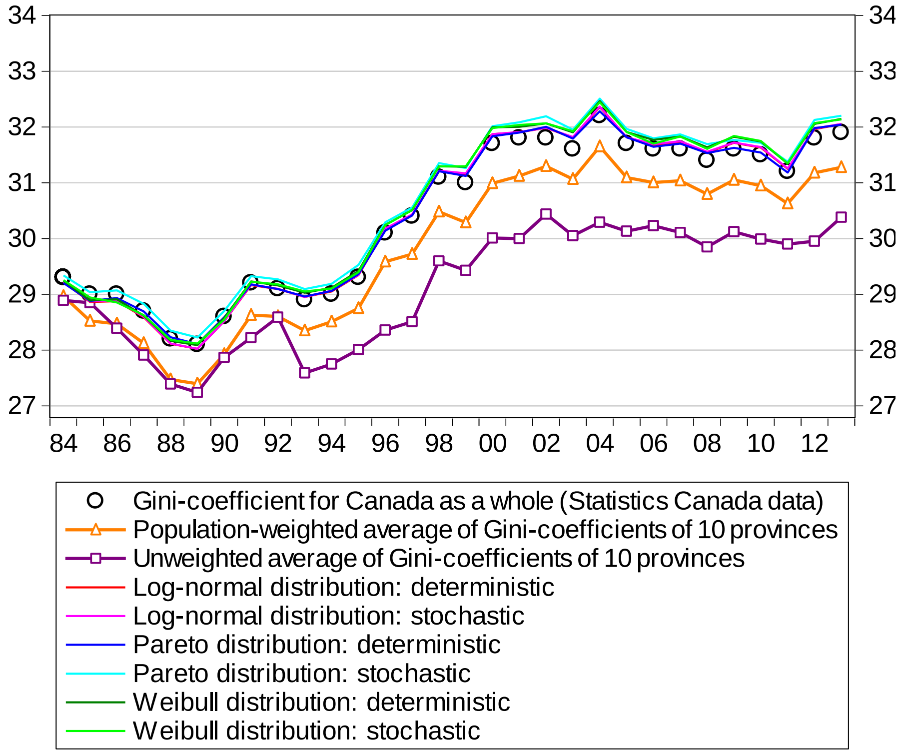
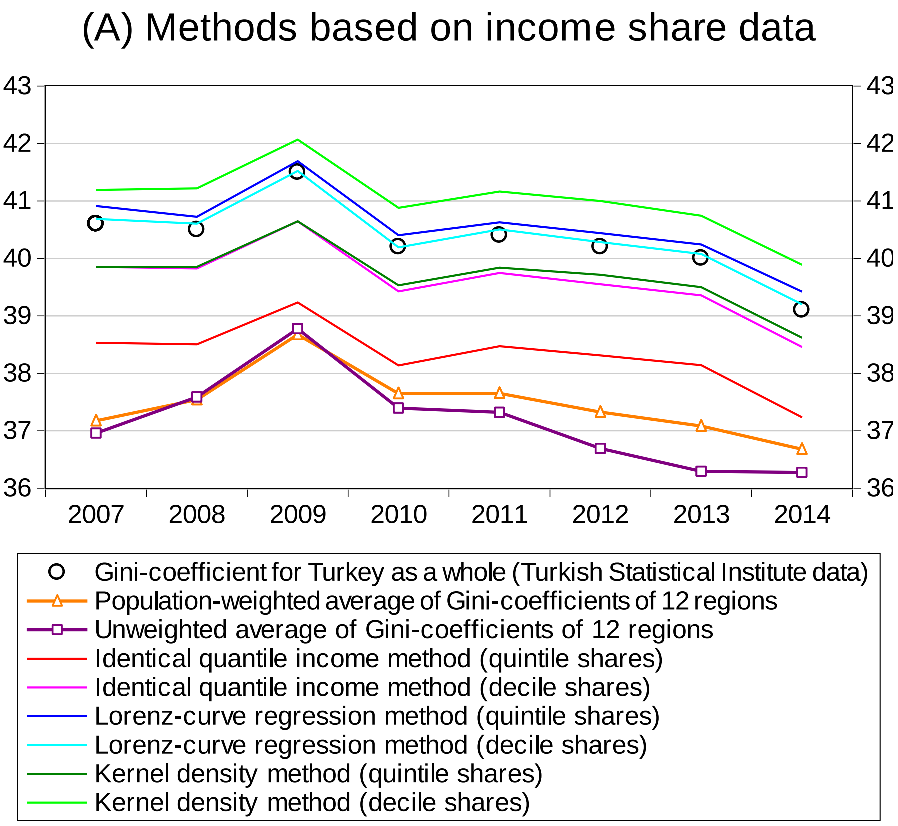


Figure 4: The overall Turkish Gini coefficient and its estimates from data of 12 regions, 2007-14



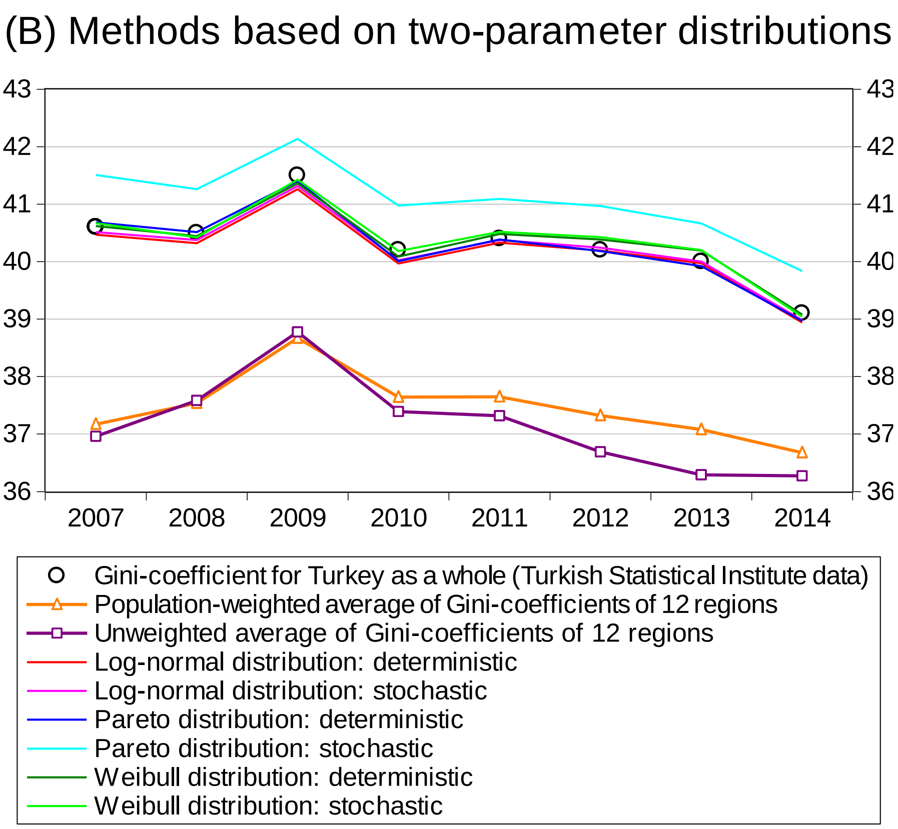


Table 3: Estimating the overall US Gini coefficient from data of 50 states and DC: average absolute difference in 2006-14

Method	Average absolute difference
Log-normal distribution (stochastic)	0.03
Lorenz-curve regression method (quintile shares)	0.03
Log-normal distribution (deterministic)	0.04
Weibull distribution (deterministic)	0.04
Weibull distribution (stochastic)	0.05
Lorenz-curve regression method (quintile and top 5% shares)	0.08
Identical quantile income method (quintile and top 5% shares)	0.14
Pareto distribution (stochastic)	0.16
Kernel density method (quintile shares)	0.20
Weighted-average state Gini	0.63
Pareto distribution (deterministic)	0.85
Unweighted-average state Gini	1.60
Identical quantile income method (quintile shares)	2.38
Kernel density method (quintile and top 5% shares)	2.57

Table 4: Estimating the overall Australian Gini coefficient from data of 8 states and territories: average absolute difference in 1995-2014

Method	Average absolute difference
Pareto distribution (stochastic)	0.26
Kernel density method (quintile shares)	0.29
Weibull distribution (stochastic)	0.32
Weibull distribution (deterministic)	0.33
Log-normal distribution (stochastic)	0.36
Lorenz-curve regression method (quintile shares)	0.39
Log-normal distribution (deterministic)	0.39
Pareto distribution (deterministic)	0.59
Weighted-average state Gini	0.74
Unweighted-average state Gini	1.53
Identical quantile income method (quintile shares)	1.86

Table 5: Estimating the overall Canadian Gini coefficient from data of 10 provinces: average absolute difference in 1984-2013

Method	Average absolute difference
Pareto distribution (deterministic)	0.08
Lognormal distribution (deterministic)	0.10
Lognormal distribution (stochastic)	0.11
Weibull distribution (stochastic)	0.16
Weibull distribution (deterministic)	0.16
Lorenz-curve regression method (decile shares)	0.20
Lorenz-curve regression method (quintile shares)	0.20
Pareto distribution (stochastic)	0.21
Identical quantile income method (decile shares)	0.25
Kernel density method (quintile shares)	0.39
Weighted-average province Gini	0.58
Identical quantile income method (quintile shares)	1.20
Unweighted-average province Gini	1.28
Kernel density method (decile shares)	1.48

Table 6: Estimating the overall Turkish Gini coefficient from data of 12 regions: average absolute difference in 2007-2014

Method	Average absolute difference
Lorenz-curve regression method (decile shares)	0.07
Pareto distribution (deterministic)	0.08
Lognormal distribution (stochastic)	0.10
Weibull distribution (deterministic)	0.10
Weibull distribution (stochastic)	0.11
Lognormal distribution (deterministic)	0.13
Lorenz-curve regression method (quintile shares)	0.24
Kernel density method (quintile shares)	0.62
Kernel density method (decile shares)	0.71
Identical quantile income method (decile shares)	0.71
Pareto distribution (stochastic)	0.74
Identical quantile income method (quintile shares)	1.99
Weighted-average region Gini	2.84
Unweighted-average region Gini	3.15

4.2 Robustness to the level of detail about quantile income shares: 27 EU and 5 non-EU European countries

We cannot carry out the aggregation test employed in section 4.1 for the entire EU because the correct overall EU-wide Gini coefficient is not available. As noted earlier, and as will be proved in section 4.3, while Eurostat publishes Gini coefficients for 28 EU members and for various groups of countries within the EU, these Gini coefficients are population-weighted averages of country-specific Gini coefficients, which are not the Gini coefficients that correspond to the combined income distribution of the countries.

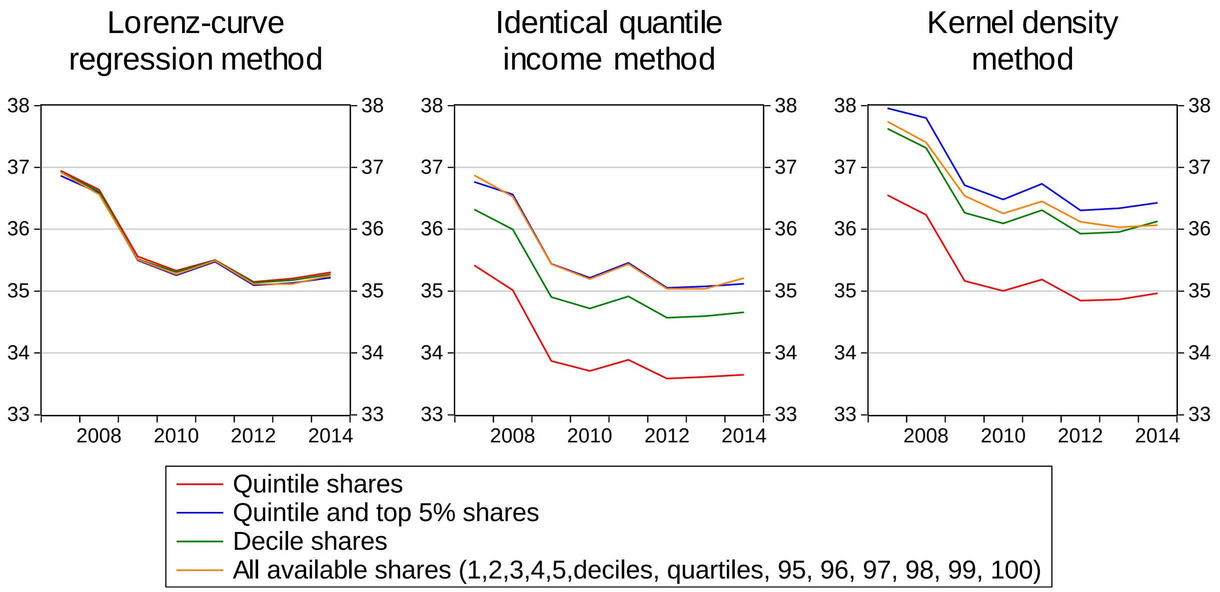
However, detailed quantile income share data is available for recent years. We therefore study the robustness of the methods relying on income share data for different levels of detail on quantile income shares used. We study four levels of detail:

1. Quintile income shares only,
2. Quintile plus top 5 percent income shares only,
3. Deciles income shares only,
4. All available income shares: 1^{st}, 2^{nd}, 3^{rd}, 4^{th} and 5^{th} percentiles, deciles, quartiles, and 95^{th}, 96^{th}, 97^{th}, 98^{th}, 99^{th} and 100^{th} percentiles.

Unfortunately, such an analysis can only be done for a relatively short period. Eurostat publishes quantile income shares data for Croatia only from 2010, Romania from 2007, Bulgaria from 2006 and most other newer EU member states from 2005. A continuous dataset for older EU member states is available also from 2005, as data for all of these countries is missing for a few or all earlier years. Therefore, calculations for the 28 members of the EU could only be made for 2010-14, for EU27 (not including Croatia) for 2007-14, and for EU25 (not including Croatia, Bulgaria and Romania) for 2005-14. Since Croatia is rather small and accounts for less than 1 percent of EU28 population while Bulgaria and Romania have a combined population share of about 5.5 percent, we decided to do the calculations for EU27 in the 2007-14 period.

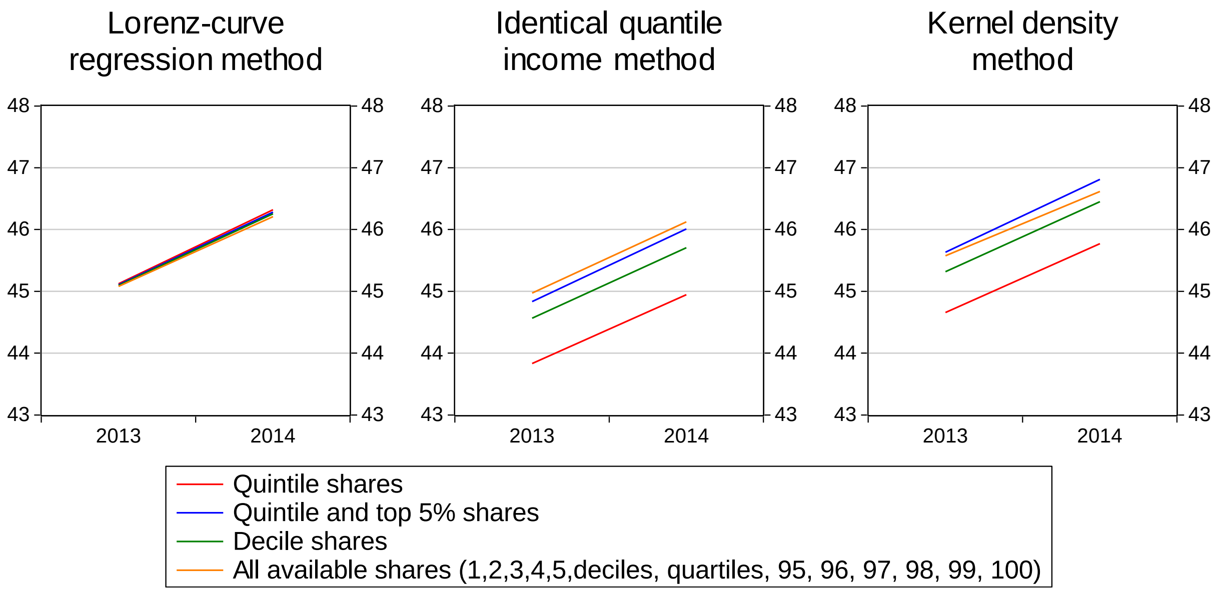
Eurostat also publishes detailed data for five non-EU countries: Iceland, Macedonia, Norway, Serbia and Switzerland. For this group of countries the same analysis can be conducted for 2013-14.

Figure 5: EU27 Gini coefficient estimates by the methods based on quantile income shares, using different levels of detail about income shares, 2007-14



Note: the 27 countries correspond to the current members of the European Union except Croatia.

Figure 6: The union of five non-EU countries' Gini coefficient estimates by the methods based on quantile income shares, using different levels of detail about income shares, 2013-14



Note: the five countries are Iceland, Macedonia, Norway, Serbia and Switzerland. Eurostat publishes detailed data on quantile income shares for these countries.

Figure 5 and Figure 6 clearly highlight the robustness of the Lorenz-curve regression method of Bhalla (2002) and Kakwani (1980): the estimates are very close to each other, independent of the level of detail regarding quantile income shares.

In contrast, the identical quantile income method of Bourguignon and Morrisson (2002) and Milanovic (2002) and the Kernel density method of Sala-i-Martin (2002) depend heavily on the level of data input detail. The identical quantile income method leads to relatively low estimates when only the quintile income share data is used – mirroring our findings for the United States, Australia, Canada and Turkey where the use of quintile shares only led to an underestimation of the national Gini coefficient. The use of decile data also leads to a somewhat lower estimate than the other estimates, while the other two data inputs (quintile plus top 5 percent share and all possible quantile shares) led to very similar results to each other as well as to the results of the Lorenz-curve regression method. This finding suggests that information about the top 5 percent income share is essential for this method, while further details may not improve the precision of this method much more.

The Kernel density method also led to substantially different results depending on the level of detail about quantile income shares. In the previous section we found for the United States, Australia, Canada and Turkey that the use of quintile income shares only has led to the most accurate results, but using more detailed income share data actually made the estimate worse. Our results for the EU27 aggregate seem to mirror this finding: when only the quintile income shares are used, the results are broadly similar to the results of the Lorenz-curve regression method and the supposedly two more accurate versions of the identical quantile income method. But when further details are used for the Kernel density method, the estimates are much higher than the results of the other methods. Results for the five non-EU countries are qualitatively the same.

4.3 Comparing the similarities of the estimates across the methods: 27 EU and 5 non-EU European countries

Figure 7 and Figure 8 compare the estimates across the methods. For each method we use only one version. For the Lorenz-curve regression method and the identical quantile income share method we report the results based on the most detailed data input on quantile income shares. For the Kernel density method, we use the results when only the quintile income shares are used. For the two-parameter distribution method we report the results based on the deterministic version. We also include the unweighted and population-weighted average of the Gini coefficients of the countries, as well as the EU27 data published by Eurostat on Figure 7.

Figure 7: Estimates of the EU27 Gini coefficient from data of 27 countries, 2007-14

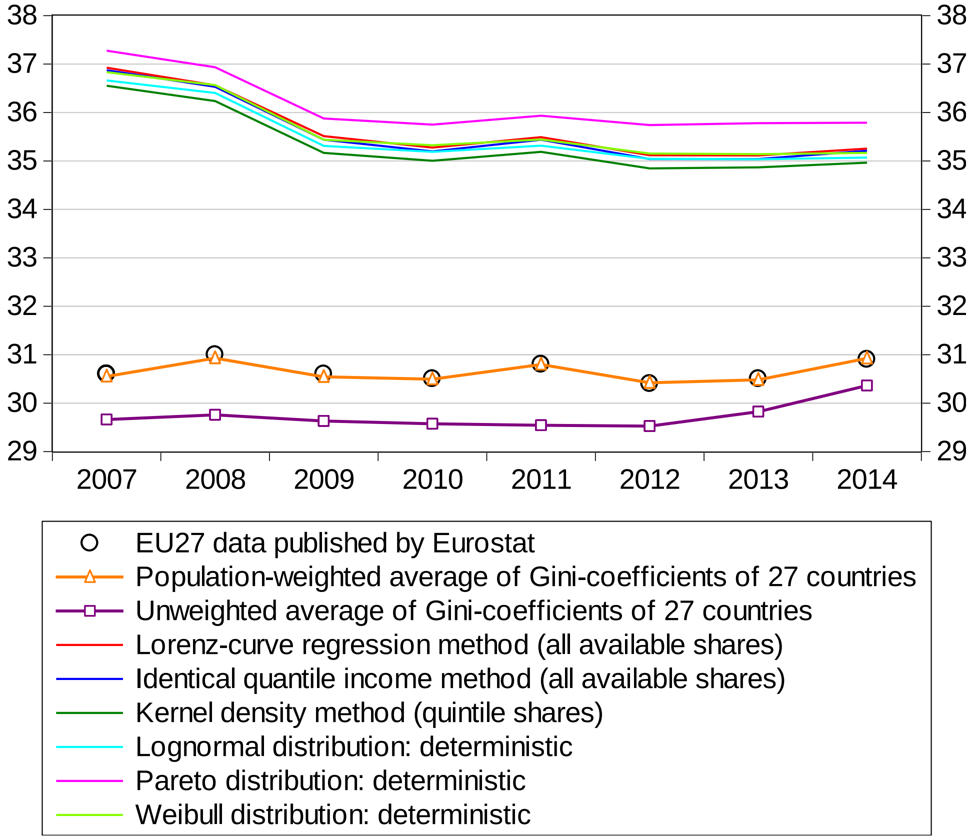


Figure 8: Estimates of the union of five non-EU countries' Gini coefficient from data of 5 countries, 2013-2014

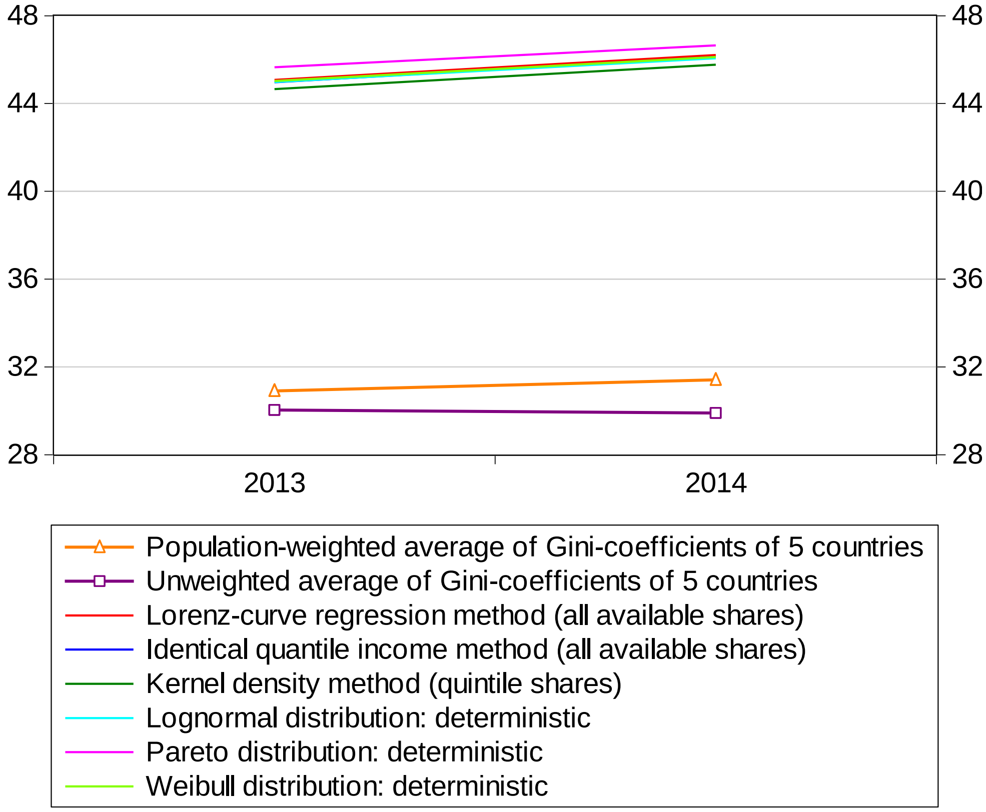


Figure 7 and Figure 8 allow us to arrive at a number of key conclusions.

First, all methods suggest that the Gini coefficient of the citizens in the union of various countries is higher than the average of country-specific Gini coefficients, thereby corroborating our conclusions from the US, Australia, Canada and Turkey in section 4.1, where we found that the country-wide Gini coefficient is higher than the average of territorial Gini coefficients.

Second, for the EU27, the data published by Eurostat is the population-weighted average of the Gini coefficients of the 27 countries and is not the Gini coefficient corresponding to the citizens living in the union of the 27 countries. We found the same results for other EU country (28, 25 and 15 countries) and euro-area Gini coefficients published by Eurostat. We therefore recommend that Eurostat stop publishing these misleading Gini coefficients for the EU and the euro-area aggregates and instead calculate the EU-wide and euro-area wide indicators of income distribution, either by combining household-level data from all countries, or by using one of the estimates presented in our paper.

Third, the results of the six methods used to calculate the EU27 Gini coefficient are very close to each other: the range of the six estimates is 0.8 Gini points on average for the EU27 and 0.9 for the five non-EU countries. The Pareto distribution has always led to the highest result: when we exclude it, the average range of the remaining five methods is only 0.3 Gini points for the EU27 and 0.4 for the five non EU countries, which are quite narrow ranges.

While this finding is based only on the calculations for two groups of countries, we hypothesise that this is a general result that could also apply to other groups of countries, not least because these findings are in line with our results obtained when calculating the country-wide Gini coefficient from territorial data for the United States, Australia, Canada and Turkey in section 4.1. As a result, we conclude that the way within-country income distribution is approximated is less important (provided, of course, that the right level of detail is used for the methods based on quantile income shares).

This finding also implies that many criticisms formulated in the literature rest on weak grounds. For example:

- Milanovic (2002) criticised the log-normal distribution approximation of Chotikapanich, Valenzuela and Rao (1997) as *"unsatisfactory"*, by arguing that income distributions cannot be well predicted from the Gini coefficient and that it is unacceptable to assume that all distributions follow a parametric pattern. Yet as we demonstrated using the estimation of US, Australian, Canadian and Turkish Gini coefficients from territorial data of these countries, the methods based on two-parameter distributions work better than the identical quantile distribution method of Milanovic (2002). For Eurostat data we found that the method of Milanovic (2002) depends a lot on the level of detail on income shares, and when (correctly) sufficiently detailed data is used, the results of his method are almost identical to the result of the two-parameter distribution methods.

- Milanovic (2003) criticised the Kernel density method of Sala-i-Martin (2006) and his results as "*very dubious*"[11], yet when the correct level of detail on the income distribution is used (at least the top 5 percent income share for the identical quantile income method, only quintile shares for the Kernel based method as in Sala-i-Martin, 2006), the methods of Milanovic (2003) and Sala-i-Martin (2006) lead to almost identical results.
- Chotikapanich, Griffiths, Rao and Valencia (2012) criticised both their earlier work using the log-normal distribution in Chotikapanich, Valenzuela and Rao (1997) by being restrictive, as well as the works of Milanovic (2002) and Sala-i-Martin (2006) for the "*untenable assumption … that persons within each income group receive the same income*". Yet we found that the log-normal distribution works extremely well in estimating US, Australian, Canadian and Turkish Gini coefficients from territorial data, while the methods of Milanovic (2002) and Sala-i-Martin (2006) also work reasonably well when the right level of detail on income shares is used[12].

5. Global and regional income inequality

Having concluded in the previous section that the two-parameter distribution method is highly reliable for estimating the Gini coefficient of income inequality for a group of countries, we use this method to calculate global and regional Gini coefficients of income inequality.

5.1 Data

The 5.1 version of the SWIID dataset includes Gini coefficients for 174 countries (some of which, such as the USSR, Yugoslavia or Czechoslovakia, do not exist anymore). Of these 174 countries, there are 59 countries with data available for each year from 1989-2013, while for 70 countries the number of missing observations was fewer than 10 in this period. We exclude Puerto Rico because of missing GDP per capita data, while for the remaining 69 countries we approximate the missing observations by assuming that the change in the Gini coefficient in the years for which data is missing was the same as the change in the simple average of Gini coefficients of countries in their region[13]. Thereby, we have a sample of 128 countries for 1989-2013. These 128 countries account for about 92 percent of global population.

Furthermore, Gini coefficients (after taxes and transfers) are available from Eurostat up to 2015 for EU countries: for these countries we use SWIID data for 1989-13 and for 2014-15 we chain Eurostat data to SWIID data. Thereby, for EU countries we can calculate the net (after taxes and transfers) Gini coefficient for 1989-2015. Eurostat does not publish data on market (ie before

[11] The working paper version of Sala-i-Martin (2006) was published in 2002 and Milanovic (2003) criticised this earlier version, which has used practically the same methodology as the 2006 journal article.
[12] We note that the criticism of the Kernel density method of Sala-i-Martin (2006) by Chotikapanich, Griffiths, Rao and Valencia (2012) is not correct at least in one aspect, because Sala-i-Martin (2006) did not assume that persons within each income group receive the same income, but he used a Kernel density method to approximate the income shares of the 100 percentiles.
[13] For this extrapolation, we grouped all developed countries into one group, while for emerging and developing countries we differentiated five groups: Asia, Africa, Central and Eastern Europe, Commonwealth of independent States and Latin America.

taxes and transfers) Gini coefficients and therefore we use only SWIID data for the EU for market inequality estimates in 1989-2013[14].

Population and GDP per capita at purchasing power parity are from the IMF World Economic Outlook database.

We keep the composition of all country groups constant throughout the sample period to avoid the impact of compositional changes on global and regional income distributions. For example, for EU28 we consider the union of the current 28 members in the full sample for 1989-2014, even though in 1989 the European Communities – the predecessor of the EU – had only 12 members.

Further details about our data sources are provided in the Annex.

[14] In addition to net (after taxes and transfers) Gini coefficient, Eurostat also publishes Gini coefficients after taxes but before transfers.

Table 7: Gini coefficient of net income inequality for 128 countries, using nine versions of the two-parameter distribution method, 1989-2013

	Lognormal			Pareto			Weibull			Average country Gini	
	Deterministic	Stochastic		Deterministic	Stochastic		Deterministic	Stochastic		Unweighted	Population-weighted
		Simple	Full		Simple	Full		Simple	Full		
1989	66.82	66.86	66.87	67.56	68.10	67.94	67.36	67.36	67.39	35.79	36.33
1990	66.91	66.93	66.95	67.67	68.08	68.12	66.81	66.82	66.89	35.91	36.78
1991	66.96	66.99	67.00	67.78	68.19	68.16	67.13	67.13	67.16	36.79	37.70
1992	66.92	66.96	66.97	67.79	68.22	68.18	67.54	67.54	67.53	37.34	38.68
1993	66.66	66.70	66.71	67.57	67.98	68.01	66.94	66.95	66.96	37.78	39.39
1994	66.50	66.53	66.54	67.43	67.90	67.86	67.18	67.20	67.04	38.01	39.68
1995	66.21	66.24	66.25	67.16	67.60	67.60	66.14	66.15	66.26	38.31	39.96
1996	65.69	65.72	65.73	66.63	67.09	67.09	66.05	66.07	65.97	38.45	39.54
1997	65.56	65.60	65.60	66.54	66.98	66.99	65.40	65.41	65.47	38.50	39.69
1998	65.56	65.60	65.60	66.50	66.96	66.94	65.55	65.56	65.59	38.65	39.66
1999	65.41	65.44	65.45	66.36	66.83	66.85	65.64	65.65	65.66	38.59	39.66
2000	65.50	65.54	65.54	66.49	66.96	67.01	65.34	65.36	65.56	38.59	40.23
2001	65.24	65.27	65.28	66.32	66.75	66.75	65.35	65.37	65.39	38.49	40.86
2002	65.20	65.24	65.24	66.39	66.86	66.96	65.43	65.44	65.45	38.43	41.92
2003	64.65	64.69	64.69	65.88	66.33	66.37	64.97	64.98	64.99	38.31	42.00
2004	64.33	64.37	64.37	65.63	66.09	66.11	64.51	64.53	64.69	38.29	42.46
2005	63.75	63.78	63.79	65.08	65.59	65.56	63.82	63.84	63.78	38.19	42.50
2006	63.21	63.25	63.26	64.58	65.05	65.10	63.46	63.47	63.39	38.05	42.65
2007	62.51	62.54	62.56	63.91	64.43	64.38	62.49	62.50	62.54	37.93	42.70
2008	61.88	61.92	61.92	63.33	63.83	63.82	61.95	61.96	61.97	37.75	42.73
2009	60.75	60.79	60.79	62.24	62.74	62.86	60.73	60.74	60.76	37.60	42.69
2010	60.22	60.25	60.27	61.72	62.21	62.22	60.34	60.36	60.37	37.51	42.68
2011	59.70	59.73	59.75	61.21	61.69	61.70	59.71	59.72	59.71	37.26	42.45
2012	59.29	59.32	59.34	60.80	61.30	61.40	59.29	59.31	59.38	37.13	42.35
2013	58.81	58.84	58.86	60.29	60.87	60.92	59.01	59.02	59.04	36.97	42.18

Source: Bruegel. Note: the mean of the 100 estimates are reported for the stochastic versions.

5.2 Global Gini coefficient estimates using nine versions of the two-parameter distribution method

From Table 7 we can draw out a number of interesting conclusions. The key methodological conclusions are the following:

- While there are some differences in the levels of the global Gini coefficients depending on the statistical distribution we use, the differences are relatively small and the dynamics are the same[15]. As regards the level, the estimates from the log-normal and Weibull distributions are very close to each other, while the use of the Pareto distribution leads to slightly higher estimates, echoing the result obtained in section 4.3.
- The mean of estimates derived from the simple and full versions of the stochastic versions are practically identical for each year when the Log-normal distribution is used (the largest yearly difference is a mere 0.02, while the average of yearly differences from 1989-2013 is 0.01). For the Pareto and Weibull distributions, the average difference from 1989-2013 is similarly small (0.01 and 0.02, respectively), while the largest yearly difference is 0.20 for the Weibull distribution and 0.16 for the Pareto distribution, which are still relatively small[16].
- Similarly to all evidence presented in earlier parts of our paper, the global Gini coefficient is greater than the average of country-specific Gini coefficients. This is most likely the result of large differences in average income among the 128 countries.
- The dynamics of the global and the country-average Gini coefficients can be different. The population-weighted average Gini coefficients of the 128 countries increased from 1989-2009, while the global Gini coefficient actually declined from 1991-2013.

The key findings regarding the level and dynamics of global income inequality:

- In most years, global inequality was higher than within-country inequality in any country[17].
- There was a slow but steady decline in global inequality from 1989-2002, since when the decline has accelerated. The recent global financial and economic crisis has not changed this trend.

Because of the similarities in the results of the nine versions of the two-parameter distribution methods we considered, for the rest of our calculations we report only the result of the deterministic version based on the log-normal distribution (except in Section 5.4 where we study the standard error of global and regional Gini coefficients).

---

[15] We also note that our estimates are broadly similar to the estimates of other works in the overlapping period as summarised in Table 1 of section 2.

[16] Note that these numerical results hold for the particular realisations of the 100 simulations we run for each version. Another set of 100 simulations might lead to different results.

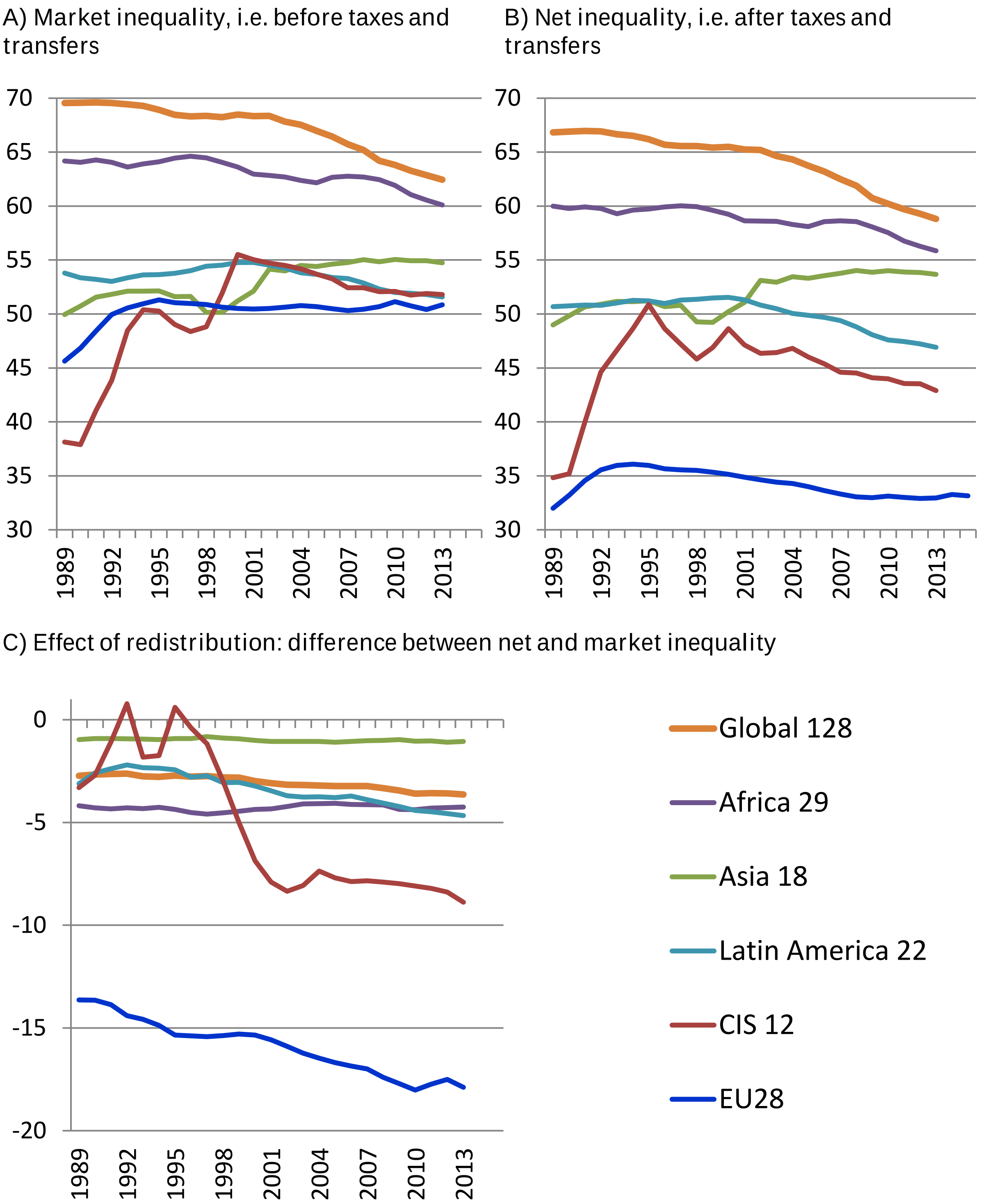
[17] Namibia is the most unequal country among the 128 countries we consider. The Gini coefficient for Namibia was larger than the global Gini coefficient in 8 years of the 1989-2013 period when using the Log-normal distribution, in 2 years when using the Pareto distribution, and 7 years when using the Weibull distribution. In all other years the global Gini coefficient was higher than within-country inequality in any country.

5.3 Regional Gini coefficients

Figure 9 compares the level and dynamics of global and regional income inequality.

- The most striking message of Figure 9 is to highlight the EU's special developments: while market income inequality in the EU has not declined and its level is similar to market inequality in other parts of the world (panel A), net inequality (after taxes and transfers) is at a much lower level in the EU than in other regions. Net inequality in the EU declined from 1994-2008, since when it has remained relatively stable (panel B). Clearly, the impact of redistribution on income inequality is much greater in Europe than in other parts of the world, as confirmed by panel C.
- Starting from much higher levels, region-wide income inequality has also declined in Africa, Latin America and the Commonwealth of Independent States (CIS – composed of most former Soviet Union countries).
- Asia is the only main part of the world where regional income inequality has actually increased, most likely because of developments in China, where within-country income inequality increased very significantly.
- The impact of redistribution on inequality is least in Asia: it reduces market inequality by a mere 1 Gini point (Panel C of Figure 9). Redistribution has small impacts (by about 4 Gini points) in Africa and Latin America. Interestingly, there was a sudden shift in the impact of redistribution from close to zero to about 7-8 Gini points in CIS countries in the late 1990s. Redistribution has clearly the strongest impact in the EU, which steadily increased during 1989-2015 to about 18 Gini points.

Figure 9: Global and regional Gini coefficients of income inequality, 1989-2013/2015

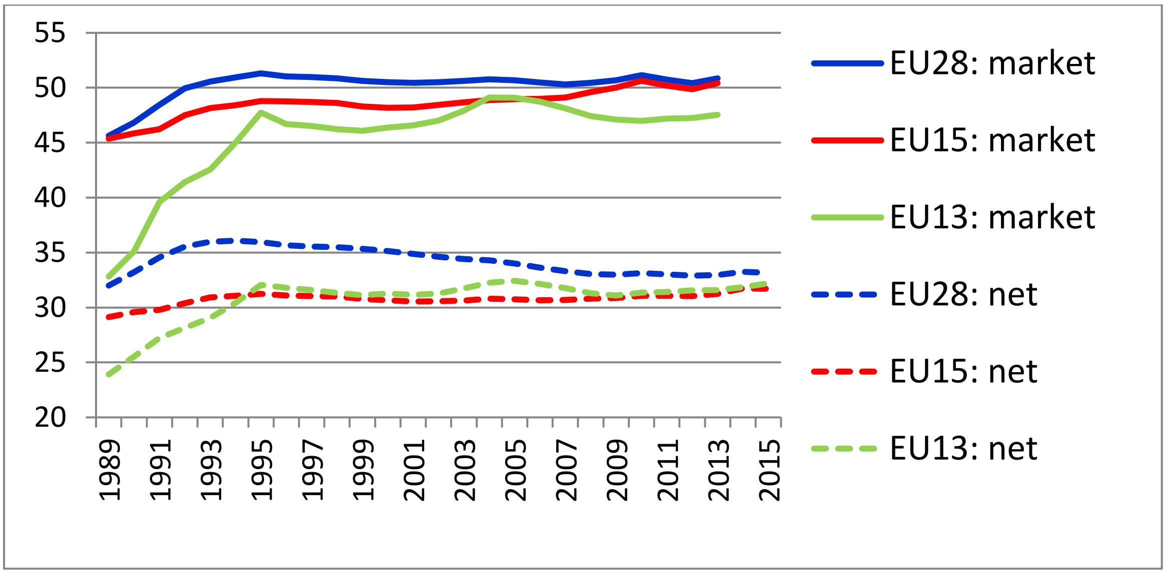


Source: Bruegel. Note: the deterministic version of the log-normal distribution is used. See the country compositions of the groups reported in the chart in the annex.

In order to obtain further insights into EU-wide inequality developments, we report results for two groups within the EU (Figure 10): the first 15 EU members[18] and the 13 EU newer members that joined between 2004-13, of which 11 countries are from central and eastern Europe (CEE).

- There was a sharp increase in EU-wide inequality between 1989-93, reflecting a moderate increase in inequality among the first 15 EU member states and a sharp increase (from a very low level) among the 13 newer member states. The CEE countries in the latter group suffered from massive output declines arising from their transition from socialist to market-based economies during this time.
- In 1995-2008 there was a sizeable decline in EU-wide net inequality, even though within-country inequality increased in many EU member states. The convergence of CEE countries (in terms of average GDP per capita) has likely played an important role in the decline of EU-wide net inequality, because within-EU15 and within-EU13 inequality remained broadly stable in this period. We will assess the role of income convergence on regional income inequality in the next section. On the other hand, since EU28 market inequality hardly changed from 1995-2013, redistribution has played a role too.
- The decline in EU-wide net inequality stopped in 2008 and since then it remained broadly stable, even though within-EU15 and within-EU13 net inequality has slightly increased in 2009-15.

Figure 10: Market and net income inequality developments in the EU, 1989-2013/2015



Source: Bruegel.

---

[18] EU15 results of Morrisson and Murtin (2004) deviate from our results by less than 1 Gini point.

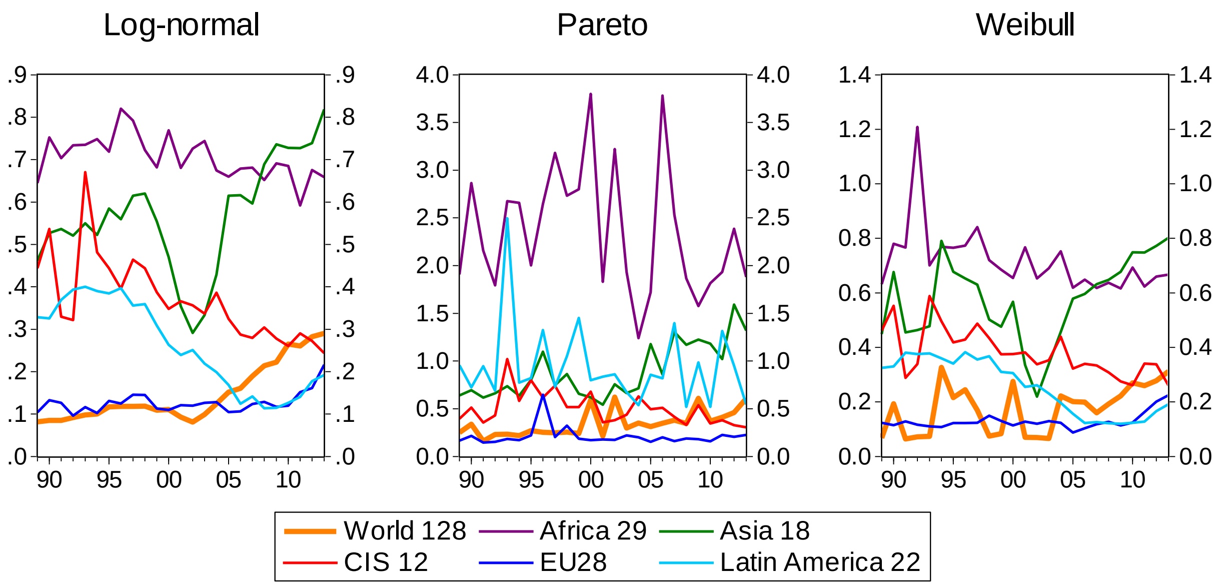
5.4 Uncertainty of global and regional Gini coefficient estimates

Using the full version of the stochastic two-parameter distribution method, we calculate the standard deviation of the 100 estimates for the global and regional Gini coefficients. Figure 11 shows that there are large differences in the estimated standard deviation depending on which distribution is used, even though the mean estimates were rather similar for all three distributions, as shown by Table 7. The use of the Pareto distribution leads to the largest standard deviation, most likely because this distribution has a fatter right tail than the other two distributions and thereby simulations from this distribution can lead to more extreme values than simulations from the other two distributions.

Figure 11 suggests that the global Gini coefficient is measured very precisely. Its estimated standard deviation has increased from about 0.1 in 1989 to about 0.3 in 2013 when using the log-normal and Weibull distributions, while it increased from about 0.3 to 0.6 when using the Pareto distribution. Such levels of standard deviation are rather small compared to the level of global Gini coefficient, which declined from about 67 in 1989 to 59 in 2013. A possible reason for the low standard deviation of the global Gini coefficient is that mean income is its key determinant and there are major differences in mean incomes in the 128 countries. Thereby the uncertainty in the country-specific Gini coefficients (which is very large in the case of some developing countries) matters less.

Among the five country groups, the standard deviation is lowest for EU countries, reflecting both the sizeable differences in mean incomes between the 28 member states and the more precise measurement of national Gini coefficients than in many developing countries. When using the log-normal and the Weibull distributions, the standard deviation of the Gini coefficient estimates for Latin America has declined to levels similar to the EU in 2007-13. On the other hand, the standard deviation of the estimated Gini coefficient is much higher in Asia and Africa.

Figure 11: Standard deviation of global and regional net Gini coefficient estimates, 1989-2013



Source: Bruegel. Note: we use the full stochastic version of the two-parameter distribution method. See the country compositions of the groups in the annex.

5.5 Decomposition of the change in global and regional Gini coefficients

Finally, we decompose the changes in global and regional net Gini coefficients to changes in within-country inequality and other factors. Unfortunately, the global and regional Gini coefficients cannot be decomposed into purely within-country and between-country inequality, unlike other indicators, like the Theil statistics (see eg Chotikapanich, Valenzuela and Rao, 1997). Therefore, using a simple numerical method, we decompose the change in global and regional Gini coefficients into four components:

- Within-country inequality,
- Mean income,
- Relative population size, and
- An 'interaction' factor, which arises from the non-linear interaction of the other three components.

To this end, first we fix the national Gini coefficients at their 1989 levels and calculate global and regional Gini coefficients using these constant national Gini coefficients and the actual values for income and population in 1989-2013. The difference between this artificial estimate and the estimate using actual data for all three key variables indicates the impact of changes in within-country inequality on changes in global and regional Gini coefficients. Second, we fix mean incomes at their 1989 levels and calculate global and regional Gini coefficients using these constant national mean incomes and the actual values of Gini coefficients and population in 1989-2013. Again, the difference between this artificial estimate and the estimate using actual data for all three key variables indicates the impact of changes in mean incomes on changes in global and regional Gini coefficients. Third, we do the same analysis with population. Finally, the difference between the actual change in global and regional Gini coefficients and the sum of the changes due to changes in within-country inequality, mean income and population indicates the interaction factor.

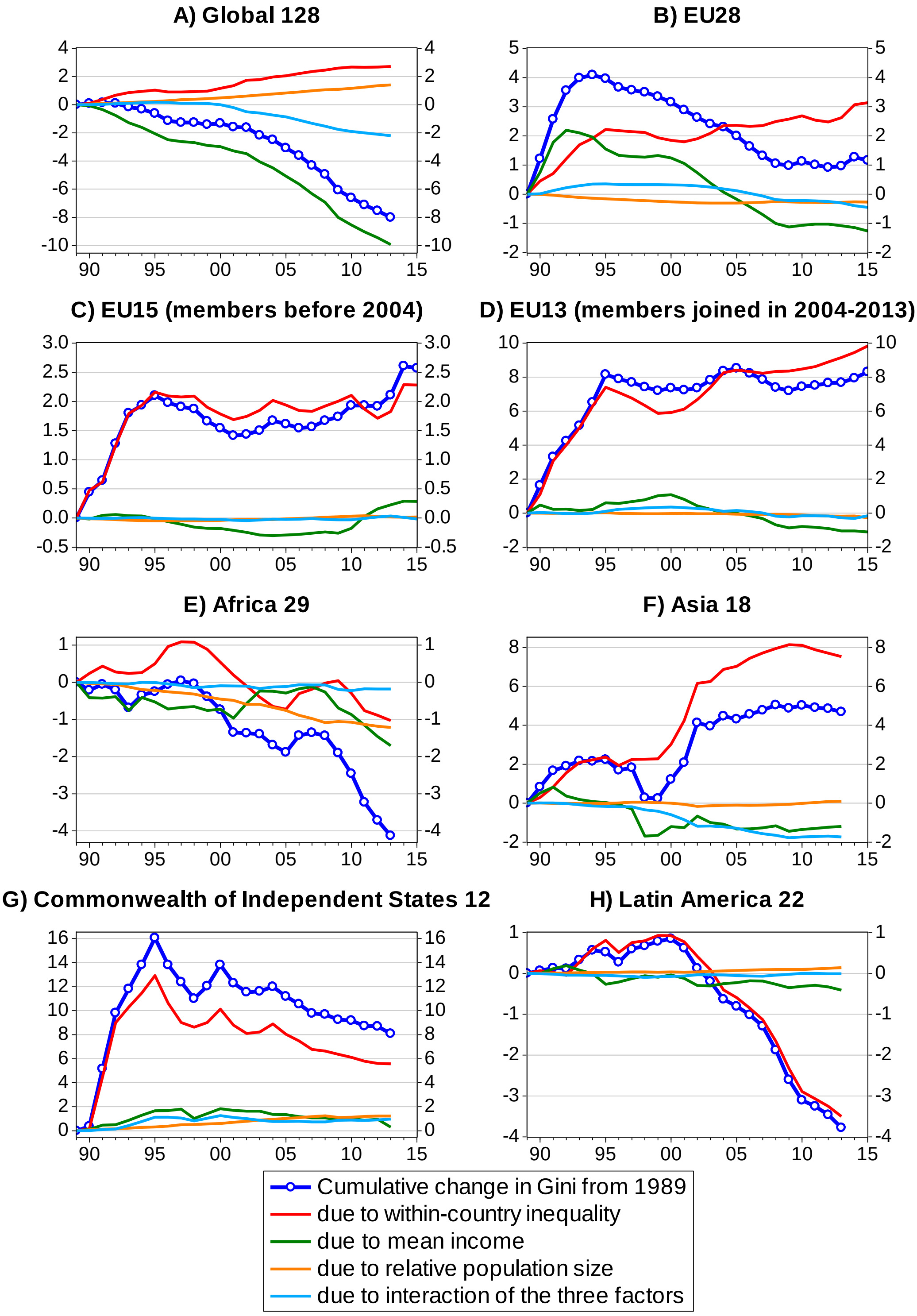
Figure 12 shows that at the global level, convergence in mean incomes was the main driving force in the reduction in global inequality and its impact accelerated in the early 2000s. While global inequality decreased by 8 Gini points (from 67 in 1989 to 59 in 2013), the convergence of mean incomes would have resulted in a 10-point decline in this period. The offsetting factors were the increase in within-country inequality, which pushed up the global Gini coefficient by about 3 points, and the change in relative population size, which increased the global Gini by about 1 point. The interactions among the three variables caused a 2 points decline.

However, in various regions of the world the relative importance of these factors varies:

- In the EU28, the 4-point increase in inequality from 1989-93 was about half-half the consequence of income divergence (collapse of CEE economies) and increases in within-country inequality. The ensuing decline in EU28 inequality in 1994-2008 was mainly the result of income convergence (minus 3 Gini points), while the increase in within-country inequality increased EU28 Gini by about 0.5 points in this period. Since 2008 the combined impact of various factors was close to zero.
- Within the two main EU groups, income convergence played a minor role and most of the change in inequality in the EU15 and EU13 was the result of within-country inequality changes. Yet in 2009-14, income divergence within the EU15 group (major economic contraction and weak recovery in some Mediterranean countries) lifted the Gini coefficient of this country group by about 0.6 Gini points.

- In Africa, the three main components had broadly the same impact.
- In Asia, the Commonwealth of Independent States and Latin America, the main driving force was within-country inequality.

Figure 12: Decomposition of the change in global and regional net Gini coefficients of income inequality, 1989-2013/2015



Source: Bruegel. Note: we use the deterministic version of the two-parameter distribution method based on the log-normal distribution. See the country compositions of the groups in the annex.

6. Summary

The goal of this paper was to estimate the global and regional distribution of income and to calculate statistics of global and regional income inequality. To this end, we first compared various methodologies that estimate within-country income distributions, which are needed to calculate the combined income distribution of citizens of different countries.

Using territorial data from the US (50 states and Washington DC), Australia (8 states and territories), Canada (10 provinces) and Turkey (12 regions) to estimate the country-level Gini coefficients, we find that the method based on simple two-parameter distributions is more accurate than three other methods using information about quantile income shares.

We also assessed the sensitivity of the quantile income shares methods to different degrees of detail about quantile income shares, using territorial data from the US, Canada and Turkey, and country-wide data from the EU. All of our calculations led to the same ranking of the three methods: the Lorenz-curve regression method of Kakwani (1980) and Bhalla (2002) is the most robust. The identical quantile income method of Bourguignon and Morrisson (2002) and Milanovic (2002) works well only if relatively detailed information is available on quantile income shares, while the Kernel density method of Sala-i-Martin (2006) works well only if quintile income data is used, but this method is less accurate when applied to more detailed income share data. When the right level of detail about quintile income shares is used, all methods work reasonably well and lead to similar results, suggesting that the way within-country income distribution is approximated is less important.

We therefore propose the use of simple two-parameter distributions to approximate within-country income distributions. This approximation is simpler, easier to implement, and relies on a more internationally-comparable dataset of national income distributions than other approaches used in the literature to calculate the global distribution of income. We found that three two-parameter distributions – the Log-normal, the Pareto and the Weibull distributions – all work well. We also suggested a simulation-based extension of the two-parameter distribution method to estimate the uncertainty in the global Gini coefficient.

We found that there was a slight decline in global inequality among the citizens of 128 countries from 1989-2002, since when the decline has accelerated. The recent global financial and economic crisis has not changed this trend. The main reason for the decline in global inequality was convergence of income per capita, which was offset to a small degree by the increase in within-country inequalities. The standard error of the global Gini coefficient is very small.

The current 28 members of the European Union are unique in terms of inequality developments. There was a sharp increase in EU-wide inequality from 1989-93, considering both market and net inequality. After 1994, market income inequality in the EU28 was at a level similar to market inequality in other parts of the world and it has remained relatively stable since then. However, net inequality (after taxes and transfers) is at a much lower level in the EU than in other regions. Moreover, net inequality in the EU declined from 1994-2008, after which it remained relatively stable. Redistribution and income convergence played major roles in the decline of EU28 net income inequality.

Regional inequality is much higher in Asia, Africa, the Commonwealth of Independent states and Latin America than in the EU28. In Asia, regional inequality has increased recent years, while it has declined in other parts of the world.

We also highlighted that the Gini coefficients for 28 EU members and for various sub-groups within the EU published by Eurostat are population-weighted averages of country-specific Gini coefficients, which are not the Gini coefficients corresponding to the combined income distribution of the countries[19]. We recommend that Eurostat stop publishing these misleading Gini coefficients for the EU and the euro area and instead calculate the EU-wide and euro-area wide indicators of income distribution either by combining household level data from all countries, or by using one of the estimates presented in our paper.

6. References

Bhalla, Surjit S. (2002) *Imagine There is No Country*, Institute for International Economics, Washington DC

Bourguignon, François and Christian Morrisson (2002) 'Inequality among World Citizens: 1820-1992', *The American Economic Review* 92(4): 727-744

Chotikapanich, Duangkamon, Rebbeca Valenzuela and D. S. Prasada Rao (1997) 'Global and Regional Inequality in the Distribution of Income: Estimation with Limited and Incomplete Data', *Empirical Economics*, XXII: 533–546

Chotikapanich, Duangkamon, William E. Griffiths, D. S. Prasada Rao and Vicar Valencia (2012) 'Global income distributions and inequality 1993 and 2000: incorporating country-level inequality modelled with Beta distributions,' *The Review of Economics and Statistics*, 94(1): 52-73

Cowell, Frank A. (2009) 'Measuring Inequality', available at: http://darp.lse.ac.uk/papersDB/Cowell_measuringinequality3.pdf

Darvas, Zsolt and Guntram B. Wolff (2016) 'An anatomy of inclusive growth in Europe', Bruegel Blueprint 26, available at: http://bruegel.org/2016/10/an-anatomy-of-inclusive-growth-in-europe/

Kakwani, Nanak C. (1980) *Income inequality and poverty. Methods of estimation and policy applications*, published for the World Bank by Oxford University Press

Lubrano, Michel (2015) 'The econometrics of inequality and poverty. Lecture 4: Lorenz curves, the Gini coefficient and parametric distributions', available at: http://www.vcharite.univ-mrs.fr/PP/lubrano/cours/Lecture-4.pdf

Milanovic, Branko (2002) 'True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone', *The Economic Journal* 112(476): 51-92

Milanovic, Branko (2003) 'The Ricardian Vice: Why Sala-i-Martin's calculations of world income inequality are wrong', available at http://econwpa.repec.org/eps/hew/papers/0305/0305003.pdf

[19] In October 2016, Eurostat published Gini coefficients for the following EU aggregates: EU28, EU27, EU25, EU15, new EU member states 12, Euro area 19, Euro area 18 and Euro area 17.

Morrisson, Christian and Fabrice Murtin (2004) ‘History and prospects of inequality among Europeans’, paper presented at the 2004 congress of the European Economic Association, available at http://www.eea-esem.com/papers/eea-esem/2004/1795/UEInequality.pdf

Sala-i-Martin, Xavier (2006) ‘The World Distribution of Income’, *The Quarterly Journal of Economics* CXXI(2): 351-397

Solt, Frederick (2016) ’The Standardized World Income Inequality Database’, *Social Science Quarterly 97*, SWIID Version 5.1, dataset available at: http://fsolt.org/swiid/

7. Annex 1: Data sources

Data sources for our main calculations for global and European Gini coefficients in sections 5 and 6 are the following:

- Gini coefficient (before taxes and transfers and after taxes and transfers): source: the Version 5.1 of the Standardised World Income Inequality Database (SWIID) of Solt (2016): http://fsolt.org/swiid/. Since data is typically available till 2013 and 2014 in the current version of the SWIID dataset, for our calculations for EU countries in the case of the Gini coefficient after taxes and transfers, we chain Eurostat data for 2014-15 to the SWIID data (ie we add to the 2013 SWIID data the change in the Gini in 2014 and 2015 as calculated from Eurostat data).
- Population: source: IMF World Economic Outlook database; we use the April 2016 version of the database: https://www.imf.org/external/pubs/ft/weo/2016/01/weodata/index.aspx.
- Mean income: similar to several other papers, and considering the arguments put forward by Chotikapanich, Griffiths, Rao and Valencia (2012), we approximate mean income with GDP per capita at purchasing power parity (PPP). Source: IMF World Economic Outlook database; we use the April 2016 version of the database (see weblink above). Some missing values for some countries at the beginning of our sample period were chained backwards to the IMF WEO data using data from World Bank World Development Indicators, European Commission's AMECO database, EBRD's Selected Economic Indicators database and the Maddison Project. We note that the IMF data is at current prices, while many researchers calculating global Gini coefficients used constant price GDP per capita from the Penn World Tables. When the goal of the analysis is, for example, the calculation of absolute poverty, such as estimating the number of people living below $2 per day, then constant price data is preferred. When the goal of the analysis is the calculation of inequality measures like the Gini coefficient or the share of population belonging to a certain quantile (as in our paper), current price data is also appropriate. Statistical offices also tend to use current price data when calculating inequality and income share indicators.

Data used in section 4.1 for the United States, Australia, Canada and Turkey are from the statistical services of these countries:

- United States: Census Bureau, http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml
- Australia: Australian Bureau of Statistics, http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6523.02013-14?OpenDocument
- Canada: Statistics Canada; income shares: http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=2060032&tabMode=dataTable&srchLan=-1&p1=-1&p2=9, Gini: http://www5.statcan.gc.ca/cansim/pick-choisir?lang=eng&p2=33&id=2060033 , population: http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=0510001&&pattern=&stByVal=1&p1=1&p2=50&tabMode=dataTable&csid= , mean income: http://www5.statcan.gc.ca/cansim/pick-choisir?lang=eng&p2=33&id=2060011

- Turkey: Turkish Statistical Institute; Gini, income shares and mean incomes: http://www.turkstat.gov.tr/PreTablo.do?alt_id=1011 ; population: https://biruni.tuik.gov.tr/medas/?kn=95&locale=en

Data used in sections 4.2 and 4.3 are from the following Eurostat datasets (http://ec.europa.eu/eurostat): Quantile shares – "Distribution of income by quantiles (source: SILC) [ilc_di01]"; Gini coefficient – "Gini coefficient of equivalised disposable income (source: SILC) [ilc_di12]"; Mean equivalised net income at purchasing power standard (PPS) – "Mean and median income by age and sex (source: SILC) [ilc_di03]".

8. Annex 2: List of the 128 countries included

country	region	country	region	country	region
Argentina	Latam22	Greece	EU28, EU15	Niger	Africa29
Armenia	CIS12	Guatemala	Latam22	Nigeria	Africa29
Australia		Guinea	Africa29	Norway	
Austria	EU28, EU15	Guinea-Bissau	Africa29	Pakistan	Asia18
Azerbaijan	CIS12	Guyana	Latam22	Panama	Latam22
Bangladesh	Asia18	Honduras	Latam22	Paraguay	Latam22
Barbados	Latam22	Hong Kong SAR		Peru	Latam22
Belarus	CIS12	Hungary	EU28, EU13	Philippines	Asia18
Belgium	EU28, EU15	Iceland		Poland	EU28, EU13
Bolivia	Latam22	India	Asia18	Portugal	EU28, EU15
Bosnia and Herzegovina		Indonesia	Asia18	Romania	EU28, EU13
Botswana	Africa29	Iran		Russia	CIS12
Brazil	Latam22	Ireland	EU28, EU15	Rwanda	Africa29
Bulgaria	EU28, EU13	Israel		Senegal	Africa29
Burkina Faso	Africa29	Italy	EU28, EU15	Sierra Leone	Africa29
Burundi	Africa29	Jamaica	Latam22	Singapore	Asia18
Cabo Verde	Africa29	Japan		Slovak Republic	EU28, EU13
Cambodia	Asia18	Jordan		Slovenia	EU28, EU13
Cameroon	Africa29	Kazakhstan	CIS12	South Africa	Africa29
Canada		Kenya	Africa29	Spain	EU28, EU15
Central African Republic	Africa29	Korea	Asia18	Sri Lanka	Asia18
Chile	Latam22	Kyrgyz Republic	CIS12	Swaziland	Africa29
China	Asia18	Lao P.D.R.	Asia18	Sweden	EU28, EU15
Colombia	Latam22	Latvia	EU28, EU13	Switzerland	
Costa Rica	Latam22	Lesotho	Africa29	Taiwan Province of China	Asia18
Côte d'Ivoire	Africa29	Lithuania	EU28, EU13	Tajikistan	CIS12
Croatia	EU28, EU13	Luxembourg	EU28, EU15	Tanzania	Africa29
Cyprus	EU28, EU13	Madagascar	Africa29	Thailand	Asia18
Czech Republic	EU28, EU13	Malawi	Africa29	Trinidad and Tobago	Latam22

Denmark	EU28, EU15	Malaysia	Asia18	Tunisia	
Dominican Republic	Latam22	Mali	Africa29	Turkey	
Ecuador	Latam22	Malta	EU28, EU13	Turkmenistan	CIS12
Egypt		Mauritania		Uganda	Africa29
El Salvador	Latam22	Mauritius	Africa29	Ukraine	CIS12
Estonia	EU28, EU13	Mexico	Latam22	United Kingdom	EU28, EU15
Ethiopia	Africa29	Moldova	CIS12	United States	
Fiji	Asia18	Mongolia	Asia18	Uruguay	Latam22
Finland	EU28, EU15	Morocco		Uzbekistan	CIS12
France	EU28, EU15	Namibia	Africa29	Venezuela	Latam22
FYR Macedonia		Nepal	Asia18	Vietnam	Asia18
Georgia	CIS12	Netherlands	EU28, EU15	Zambia	Africa29
Germany	EU28, EU15	New Zealand		Zimbabwe	Africa29
Ghana	Africa29	Nicaragua	Latam22		

bruegel

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org