Decomposing global inequality

Jørgen Modalsli*

December 23, 2015

Abstract

This paper provides an intuitive additive decomposition of the global income Gini coefficient with respect to differences within and between countries.

In 2005, nearly half the total global income inequality is due to income differences between Europeans and North Americans on the one side and inhabitants of Asia on the other, with the China-USA income differences alone accounting for six per cent of global inequality. Historically, income differences between Asia and Europe have driven a large part of global inequality, but the quantitative importance of within-Asia income inequality has increased substantially since 1950.

Keywords: Global inequality, Gini coefficient, inequality decomposition
JEL codes: D31, O15

*Research Department, Statistics Norway, P.O. Box 8131 Dep, N-0030 Oslo. Email: mod@ssb.no. Phone: +47 21094312. Fax: +47 21090040. I am grateful to Rolf Aaberge, two anonymous referees and participants at the 2012 Canazei Winter School of Inequality and Social Welfare Theory for helpful comments and discussions. Support from the Research Council of Norway is acknowledged.

1 Introduction

Global income inequality is frequently reported in terms of estimates of the Gini coefficient. The global Gini coefficient among individuals accounts for within-country as well as between-country differentials. However, because of the non-linear nature of the Gini coefficient, decomposition into within- and between-country inequality is not straightforward. This paper provides an alternative justification of the decomposition method proposed by Ebert (2010), and extends it to the case of more than two subgroups. This makes possible a more intuitive re-interpretation of existing estimates of global inequality, allocating "contributions" to global inequality to specific countries or country groups.

De-composing the Gini coefficient

An estimate of the Gini coefficient for the entire world consists of one number. But how should we think about this number, other than marking off the level of inequality on a scale between zero and one? When estimates differ, where does the difference come from? If global inequality is increasing, what drives the increase? To answer these question, we need an approach to the *decomposition* of the aggregate inequality measure.[1]

Shorrocks (1984) described formal requirements for decomposition of inequality indices such as the Gini coefficient, and defined "weak decomposition" as a structure where within- and between-group measures do not simply aggregate, but are re-weighted according to group means and sizes before the groups are added together. Later literature have mainly focused on *linear* decompositions, where the number of terms is a linear function of the number of groups; for example, Mookherjee and Shorrocks (1982) and Lambert and Aronson (1993) see global inequality in a set of S groups to be composed of S within-group terms, one between-group term and one residual term.[2]

Ebert (2010) extends and concretizes the notion of decomposability by explicitly viewing the Gini coefficient (and other related measures) as sums of all possible differences between individuals. While this interpretation of the Gini coefficient goes back to the original paper by Corrado Gini, it does not appear to have been used in this way in group decomposition before. In a two-group population, Ebert defines between-group inequality as a function of all income

[1] In this paper the term **global** will be used whenever the discussion concerns a population built up of several groups. That is, we have a set of *individuals* who are members of *groups* who together constitute the *global* population.

[2] There are, however, some studies based on a set of "overlapping" terms describing the overlap between the income distribution of the subgroups (Yitzhaki and Lerman, 1991; Yitzhaki, 1994) as well as an interpretation of between-group differences as potential gains (Pyatt, 1976). We return to these approaches in the next section.

comparisons between individuals in group 1 and individuals in group 2.[3] Ebert's decomposition can be straightforwardly generalized to more than two groups. This will be outlined in detail in the next section. The key innovation is that each between-group term is a function of income differences between individuals in two specific groups, giving a set of terms that aggregate to form the aggregate Gini coefficient. As the Gini coefficient is a function of individual income comparisons, a linear decomposition is not mathematically feasible.

In the setting of this paper, studying global inequality, we can then discuss the contributions from country and country-pairs as well as regions and region-pairs to global inequality. Using this version of Ebert's weak decomposition, we end up with a set of sub-indices that *do* add up to form the global Gini coefficient.

2 An additive subgroup decomposition

This section outlines the non-linear inequality decomposition in detail. The decomposition method is based on Ebert (2010), who defines a family of weakly decomposable measures. In this paper, we are only concerned with the Gini coefficient which is one member of this family. As stated above, we build on Ebert's between-group term (for a population with two groups) defined as the weighted sum of all income differences between individuals in group 1 and individuals in group 2. This means that the Gini coefficient can be decomposed into within-group terms for the two groups as well as the one between-group term. The following paragraphs explain the justification for such a decomposition, generalizes it to the case of more than two groups, and adds some economic intuition to the analysis of inequality within and between groups.

2.1 Setup

Consider a population of N individuals with incomes given by the income vector $\mathbf{y}$. The population is divided into S mutually exclusive groups, where the size of group s is N_s, $s = 1, 2, ..., S$. Incomes are denoted $y_{s,i}$ where s indexes groups and i indexes individuals within groups. The income vector is sorted by group membership, and can be written as

$$\mathbf{y} = \{y_{1,1}, y_{1,2}, ..., y_{1,N_1}, y_{2,1}, y_{2,2}, ..., y_{s,N_s}\} \tag{1}$$

[3] This should not be confused with the "between-group" coefficient of Lambert and Aronson, which is simply a function of differences between income means.

The relative size of group s is $p_s = N_s/N$. The mean income of group s is $\mu_s = \frac{1}{N_s}\sum_{i=1}^{N_s} y_{s,i}$, and the aggregate mean is $\mu = \frac{1}{N}\sum_{q=1}^{S}\sum_{i=1}^{N_q} y_{q,i} = \sum_{q=1}^{S} p_q\mu_q$. The Gini coefficient for the entire population is given by a scaled sum of all pairwise income comparisons

$$G = 100 \cdot \frac{1}{2N^2\mu}\sum_{q=1}^{S}\sum_{i=1}^{N_q}\sum_{r=1}^{S}\sum_{j=1}^{N_r}|y_i - y_j| \tag{2}$$

We further decompose (2) into between-group components. To make the discussion clearer, a specific example will be used in the presentation of the decomposition.

2.2 Subgroup decomposition: Example

Consider a population of seven individuals partitioned into three groups, with the income vector

$$\mathbf{y} = \{\underbrace{2, 5, 8}_{s=1}, \underbrace{5, 11}_{s=2}, \underbrace{4, 7}_{s=3}\} \tag{3}$$

For 7 individuals, there are $\binom{7}{2} = 21$ unique comparisons (not counting the self-comparisons, which will always be zero). We can lay them out as shown in Table 1.

[Table 1 about here.]

The Gini coefficient for the entire population is found by summing all the differences in Table 1, dividing by the square of the number of observations and the population mean.[4] It is evident from Table 1 that the Gini coefficient can be decomposed into within-group and between-group components. The total set of differences are either between individuals within groups, in the three diagonal boxes, or between individuals of different groups, in the three boxes at bottom left. Denoting the sum of the numbers in each of the boxes as H_{qr}, we have

[4] We could, alternatively, write differences in *all* the cells and then divide by 2, as done in Equation (2). For the purpose of this decomposition, though, it is more convenient to have each difference occur only once.

$$H_{qq} = \sum_{i=1}^{N_q} \sum_{j=i}^{N_q} |y_j - y_i| \qquad \text{when } q = r \text{ (within groups)} \tag{4}$$

$$H_{qr} = \sum_{i=1}^{N_q} \sum_{j=1}^{N_r} |y_j - y_i| \qquad \text{when } q \neq r \text{ (between groups)} \tag{5}$$

In our example, the sums of groups differences are given in Table 2.

[Table 2 about here.]

From Table 2 we get the total sum

$$H = \sum_{q=1}^{S} \sum_{r=q}^{S} H_{qr} = 74 \tag{6}$$

Dividing H by the population mean and the square of the population size gives the conventional population Gini coefficient

$$G = 100 \cdot \frac{H}{N^2 \mu} = 100 \cdot \frac{74}{6 \cdot 7 \cdot 7} = 25.2 \tag{7}$$

Each of the cells of Table 2 provides a "within" (the cells where $q = r$; Equation 4) or a "between" ($q \neq r$; Equation 5) contribution to the Gini coefficient.

We then scale these cells by the same deflator as in Equation (7) to obtain the Gini components G_{qr}

$$G_{qr} = 100 \cdot \frac{1}{N^2 \mu} H_{qr} \tag{8}$$

These components are shown in Table 3. The numbers sum to 25.2, the aggregate Gini coefficient.[5] The Table also illustrates the "contribution" of a given group q to the term G_{qr}, with the terms relating to Groups 1, 2 and 3 in the example being highlighted in the first, second and third panel of the figure. Any component of between-group inequality (the off-diagonal cells) belong to two groups, while the within-group terms on the diagonal are only affected by changes in dispersion in one group.

[5]Similar tabulations were used for (inferred) inequality decomposition in Modalsli (2015).

$$G = \sum_{q=1}^{S} \sum_{r=q}^{S} G_{qr} \tag{9}$$

[Table 3 about here.]

We have now arrived at a full decomposition of the Gini coefficient as a set of scaled sums of income differences between individuals within a given group ("within" terms, obtained when $q = r$ in Equation 8) and between individuals in two different groups ("between" terms, obtained when $q \neq r$ in Equation 8). In a population of S groups, there are S "within" terms and $S(S-1)/2$ "between" terms. While this presents a complete classification of all income differences from Equation (2) into one and only one cell, it is sometimes useful to also consider how much of a given between-group term that can be attributed to differences in *mean incomes* between two groups. This is the topic of the next section.

2.3 Group mean differences and group overlap

At this point, it is useful to compare the decomposition described here to that used in Lambert and Aronson (1993), Equation (1) (L superscripts added):

$$G = G_B^L + \sum a_q^L G_q^L + R^L \tag{10}$$

The within-group coefficients correspond directly to Lambert and Aronson's scaled within terms, $G_{qq} = a_q^L G_q^L$. The income comparisons constituting the between-group terms G_{qr}, however, are present both in Lambert and Aronson's between-group term G_B^L an the residual term R^L. It is straightforward to separate G_{qr} into a mean-between and and residual term. Such a separation highlights how much of a given income difference between individuals in two groups are due to differences in mean incomes. For two groups q and r we hence have a pair-specific between-group Gini component that is equivalent to the difference in mean incomes between the groups, scaled by the group size:[6]

$$G_{qr}^m = 100 \cdot \frac{1}{N^2 \mu} N_q N_r \left| \mu_q - \mu_r \right| \tag{11}$$

Similarly, we can define a "residual" inequality, given as $G_{qr}^r = G_{qr} - G_{qr}^m$.

[6] Such weighted differences in country means are denoted "intercountry terms" by Milanovic (2005, p. 88-89).

The "within" component is then a special case of the residual; within groups, the mean income is the same, and so all inequality is "residual" — that is, $G_{qr}^{r} = G_{qr}$ if $q = r$.

Applied to Table 3, the values in the off-diagonal cells G_{qr} can thus be split up into two components: the mean-between G_{qr}^{m} and the residual G_{qr}^{r}. This is illustrated in Table 4.

[Table 4 about here.]

If the income ranges of two different groups do not overlap, the residual term for that group interaction is zero — inequality between the means perfectly summarizes the total distance between individuals in the two distributions.

2.4 What is the inequality within group q, and between group q and r?

The within-group Gini coefficient of group q is defined as the coefficient we would get if the group was a separate population. We see that this is a scaled form of G_{qq} in (8):

$$G_{q}^{w} = \frac{1}{(p_q)^2 \left(\frac{\mu_q}{\mu}\right)} G_{qq} \tag{12}$$

and it will be convenient to similarly define scaled between-group Gini coefficients as

$$G_{qr}^{w} = \frac{1}{p_q p_r \left(\frac{\mu_q}{\mu} + \frac{\mu_r}{\mu}\right)} G_{qr} \tag{13}$$

The particular scaling in (13) — using twice the unweighted arithmetic mean of group mean incomes — merits further explanation. One could think that the most intuitive approach would be to weight these means by group sizes. However, for each comparison of incomes in H_{qr}, there is exactly one individual from group q and one from group r — fifty percent of each. As we sum all the comparisons, this ratio holds. Moreover, this scaling ensures that full between-group inequality is 1, giving a similar interpretation to the within-group inequality.

There are two ways in which scaled between-group inequality could be 1. First, all individuals in group q could be extremely rich, while all individuals in group r had zero income. As income in the rich group approaches infinity, we have $G_{qr}^{b} \to 1$. In that case, the inequality is purely driven by the mean-difference component as defined in Section 2.3. The other way we could have "complete inequality" would be to have one agent in each group holding all the wealth of that group, equal for both groups, with the rest of the individuals having zero income. As the mass of those two agents both approach zero, we have $G^b_{qr} \to 1$, and as group means are equal, inequality is entirely driven by the residual term (distributions overlap perfectly).

The group-scaled measures in (12) and (13) have different uses than the globally scaled measure in (8). Scaled within-group measures (G^w) can be used to compare inequality in different sub-populations; seeing which is "more" unequal by this particular inequality measure. Similarly, the scaled between-group measure G^b can be used to assess the distance between the income distributions of two populations, compared to a hypothetical maximum and minimum.

The measures (8), on the other hand, are scaled by aggregate population size and mean income, weight all individuals in the population equally and are well-suited to asking questions about the aggregate population: what contributes to overall inequality? This, for the specific application of global income inequality, will be answered in the next section.

2.5 Comparisons to other decomposition methods

The decomposition of inequality into within- and between group terms is presented here as an extension of Ebert (2010). It shares with the alternative approaches mentioned in the Introduction (footnote 2) a focus on the full set of group comparisons. However, while the present approach focuses on categorizing the between-individual differences that make up the Gini coefficient into predetermined groups, the other methods have different justifications.

Pyatt (1976) presents a matrix $\mathbf{E}$ where an element in row i, column j is the expected gain for a random individual in group i from a choice between the individual's own income and that of a random individual from group j. Hence, Pyatt's $\mathbf{E}$ has S^2 terms in contrast to the $S^2/2 + S$ terms presented in Table 3, with the highest cell values obtained when the mean income of group i is lower than that of group j. Table 3 can be obtained from a $\mathbf{E}$ matrix by summing the ij and ji terms and dividing each cell by the square of the group population size.

Yitzhaki (1994), extending an approach proposed by Yitzhaki and Lerman (1991), constructs a decomposition of the Gini coefficient with the aim of assessing how grouping of individuals reflect different layers in the distribution being studied. Yitzhaki's "overlapping index" reflects to what extent the income distribution of one group is contained in the income distribution of another group. The S^2 overlapping indices O_{ji} can be further collapsed into S indices O_i, each indicating the overlapping of the distribution of a given group with respect to the overall distribution. The between-group decomposition of Yitzhaki (1994) is based on the mean rank of group members rather than mean income as in Lambert and Aronson (1993). These differences in purpose makes a direct comparison of the results difficult. We will, however, return to applications of Yitzhaki's decomposition to global income inequality in Section 4.

3 Global inequality

With the allocation of individual income differences to group pairs presented in the previous section, we now turn to the implementation of this decomposition method to global inequality. To estimate income inequality for the world as a whole, one has to construct a global income distribution based on within-country income dispersion data. The research on global inequality up to 2006 is summarized in Anand and Segal (2008), who give tables of the estimates and thorough discussion on several methodological issues.[7]

The data on the global income distribution used in the present paper is for 2005 and was obtained from country-specific nominal decile mean incomes collected by Milanovic (2010). The data covers a total of 117 countries comprising 93 per cent of world population. Country PPP conversion rates are obtained from the comparisons conducted at the World Bank (the International Comparison Program, ICP).[8]

An objective of this paper is to allocate all global interpersonal income differences (together composing the global Gini) to a specific between- or within-term, both at the country and (continental) region level. As the decomposition depends on population sizes, it is not desirable that regions with lower coverage (which also tend to be poorer) get lower weights in the decomposition of global inequality. For this reason, the income distributions for the countries

[7]The first globally comprehensive attempt in estimating global inequality on the basis of country-specific distribution data was Bourguignon and Morrisson (2002), who showed a steadily increasing global Gini coefficient from 1820 to 1992, reaching 0.66 in 1992. Earlier studies use very restricted subsets of world countries. An exception is Chotikapanich et al. (1997) have a comprehensive country coverage, but assume log-normal distributions within all countries and back out the dispersion parameter from published country Gini coefficients. Atkinson and Brandolini (2001) give a review of the early literature comparing inequality across countries and caution against "mechanical" use of databases of country characteristics, such as pre-calculated Gini coefficients.

[8]Much of the variation in estimates of the global Gini coefficient comes from different uses of PPP measures. Dikhanov and Ward (2001) and Dowrick and Akmal (2005) construct their own PPP measures and find higher levels of global inequality in the early 1990s. The International Comparison Programme (ICP), initiated by the World Bank, led to a substantial revision of assumed price levels in different countries; the implications for global inequality are outlined in Chen and Ravallion (2010) and Milanovic (2010). In short, the main effect of the ICP adjustment is that price levels in several important poor countries are adjusted up, leading to higher measured inequality between individuals in low-income countries and individuals in high-income countries. The new estimates in Milanovic (2010) are higher than those following previous price level adjustments, giving a Gini coefficient of 71 in 2002.

with missing data were imputed from earlier income distributions or from neighboring countries, bringing the number of countries up to 188 (see Appendix for detail). This imputation gives a global Gini coefficient of 69.7 rather than the 70.7 reported by Milanovic.

3.1 Country contributions to global inequality

We start by considering the decomposition of global inequality by country. With 188 countries in the sample, there are a total of $189 \cdot 188/2 = 17766$ terms, each of which consists of a country pair (income differences between individuals in two separate countries) or one country (the within-country Gini coefficient, scaled by population and income size). Two of these terms account for 10% of global income inequality; the 78 largest of the 17766 terms for 50%, and we would have to examine 1400 to get to 90% of global inequality. In this subsection we will restrict our attention to the twenty largest terms, which are presented in Table 5. These terms account for a total Gini contribution of 21.9 or almost one third of the global Gini coefficient.

[Table 5 about here.]

The country pairs at the top of the list all have large populations and/or big income differences. The China-USA interaction contains nearly one hundredth of all potential individual comparisons in the world, and most of these comparisons give large income differences. In total, the China-USA comparison can "explain" six per cent of the global Gini coefficient, of which nearly all is contained in a simple comparison of the mean incomes of China and USA.

It should be noted here that as the comparisons are based on surveys and further simplified to decile or vintile data, the upper tails of the income distributions are not adequately represented. Milanovic (2010) show that this has only a small impact on the countrywide Gini coefficient. Nearly all this impact, however, is likely to be on the residual term, which is therefore underrepresented in the table above.

Two within-country inequality terms appear on the list; China in position seven and the United States in position twenty. The smallest countries (by population) to appear on the list are the United Kingdom and Italy (both have high incomes); the lowest-income countries to appear are Bangladesh and Nigeria (both with large populations).

It is evident that the contribution of income differences between China, India and USA contribute substantially to the world inequality; the three "between" terms in position 1, 2 and 4 sum to 13.5% of the global Gini coefficient. The differences in mean income between these three countries are denoted "the triangle that matters" by Milanovic (2005, p. 88f). By constructing "triangles" from the dyads in a full 17,766-row version of Table 5 we examine whether any other such triangles have high quantitative importance. There are a total of $188 \cdot 187 \cdot 186/6$ such triangles, or slightly more than one million. All the quantitatively most important triangles involve USA and either China or India, combined with a third country. The triangle with the largest contribution that involves neither USA, China or India is Brazil-Indonesia-Japan; income differences between these three countries account for only 0.8 per cent of global inequality.

We can further utilize the scaled versions of the Gini contributions from Section 2.4 to assess to what extent the contribution of a given country pair follows from the size of the country pair (in terms of mean income or population size) or from proportionately large income differences given these sizes. These coefficients are shown in the rightmost column of Table 5. The China-USA term $G_{qr} = 4.2$ is the product of a large between-country scaled Gini $G_{qr}^{w} = 81.6$, an income weight of 5.4 (reflecting in particular the high mean income of the United States) and a population weight of one per cent. The India-USA term, on the other hand, has a lower population weight but a higher scaled between-group Gini. In general, the distribution of between-country scaled Gini terms is much more dispersed than the within-country Gini coefficients. While the 10th, 50th and 90th percentile within-country Gini are 30, 39 and 55, respectively, the similar distribution for the scaled between-country terms are 41, 62 and 90. This reflects in particular the large contribution from between-country mean income differences to the between-country terms.

While a study of the quantitatively most important country pairs adds to our understanding of how to interpret the global Gini, the large number of terms in a by-country decomposition prohibits a full account of all global income differences. For this reason, we also consider a decomposition into aggregate regions, where all terms can be listed.

3.2 Regional inequality

To construct regions for the purpose of a decomposition of global income inequality, countries must be grouped together into larger units. To this purpose, we start with the United Nations "geoscheme" dividing the world into six regions: Africa, Europe, Latin America and the Caribbean, Northern America, Asia, Europe and Oceania.[9] As Asia comprises more than sixty per cent of world population, it is desirable to split this region into at least two components. The next level down in the UN scheme is the 22 "sub-regions", five of which are in

[9] "Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings", available at http://unstats.un.org/unsd/methods/m49/m49regin.htm

Asia. Of all possible groupings of these Asian regions, the one with most similar population sizes and a contiguous geographic grouping is to group the regions of East Asia and Southeast Asia together, with the remaining region consisting of West Asia, South Asia and Central Asia. This gives a total of seven regions, the key properties of which are listed in Table 6.

[Table 6 about here.]

[Table 7 about here.]

The income distribution for each region is constructed from the country distributions in Milanovic (2010). The Gini coefficient is decomposed into one number for the inequality within a region and one number for each region pair. These components, leading to the world Gini of 69.7, are given in Table 7 and illustrated in Figure 1.

Between-region inequality

As is clear from the table, most of the world's inequality comes from the difference between high- and medium-income regions, and in particular the differences between Europe, North America and Asia. All cells with Gini contributions larger than 4.0 involve one of the two Asian regions and/or Europe. The combination of high population in the Asian regions, giving high population weights, and high mean income in the European region, giving high average income distances, mean that together, interactions between and within these four regions constitute a Gini contribution of 47, or two thirds of total global inequality.

[Table 8 about here.]

We can further explore the contributions of these regions by disaggregating the cells in Table 7 into between and residual components, as discussed in Section 2.3. This is shown in Table 8. We see that there is little overlap between the Asian regions on the one side and Northern America and Europe on the other. The largest residual component of the four interaction terms is 1.6 for the interaction between East/Southeast Asia and Europe, which reflects the fact that parts of Europe is middle-income (Russia and parts of East Europe) while parts of East Asia is high-income.

The Gini contribution of the interaction between the two parts of Asia defined here is 7.1, of which 4.6 comes from the income of East/Southeast Asia being higher than South/West/Central Asia. Economically, the East/Southeast Asia region is very diverse, with a contribution from within-region differences of 5.1, or around seven per cent of the global Gini coefficient. If we consider

Asia as a whole, we add the interaction term to the two within-region terms to get a total within-Asia contribution to global Gini of 1.5 + 5.1 + 7.1 = 13.7, or almost twenty per cent of the global Gini coefficient.

The largest cells in Table 7 not involving a citizen of Asia in at least one in the terms of Equation (2) are the interactions between Africa on one side and Europe and Northern America on the other, at 3.4 and 3.2, respectively. These are almost fully driven by differences in mean incomes; the African income distribution hardly overlaps with either the European or North American one. There are also sizable terms from the comparison of Europe and North America to Latin America; however, in this case, the residual term does account for some of the difference. Oceania, with its very low population size, contributes little to global inequality.

Within-region inequality

Within-group inequality, presented in the diagonal of Table 7, accounts for a contribution of 9.2, or 13 per cent of global inequality. As mentioned above, East and Southeast Asia, which contains both Japan, Taiwan and other rich countries combined with middle-income countries such as China and Indonesia, contributes most of this. Inequality between Russia and Western Europe is part of the reason while the European within-cell is much higher than the Northern American one.

It might be surprising that within-region inequality in Latin America is very low, at 0.3. The same goes for Africa, at 0.4. Latin America is a continent of large differences, and the scaled within-region Gini coefficient is indeed quite high. However, compared to other regions of the world, Latin America is not very big; less than one tenth of the world's population live there. Hence, income comparisons between Latin Americans constitute less than one hundredth of all income comparisons in the world. China alone has twice the population, and hence four times the income comparisons, that Latin America has. Hence, even though Latin America's relatively high mean income corresponds to large within-region (and in particular within-country) differences, this is not a large part of total world inequality.

[Figure 1 about here.]

3.3 Inequality in regions and between region pairs

In the discussion so far, all the Gini contributions are scaled to global means and population sizes; this is useful because we get a clearly identified contribution to world inequality. However, these numbers do not inform us about how unequal regions or region-pairs are compared to the highest possible inequality. To compare inequality within and between regions on the basis of region means and population sizes, we use the scaled inequality measures discussed in Section 2.4. As noted there, the within-region components correspond to what are usually called "group Ginis"; inequality that we would get if each region was a separate population. Scaled inequality for the seven regions is given in Table 9.

[Table 9 about here.]

The highest within-region Gini is found in the two Asian regions. This is not surprising, as both group high- and low-income countries together. Inequality within Africa and Latin America is at an intermediate level, while there is relatively low within-region inequality in Northern America, Europe and Oceania.

All the within-region Gini coefficients are substantially lower than world inequality, while at the upper range of the world's within-country Gini coefficients. This can be expected as the grouping removes some of the large global heterogeneities while still grouping together countries with very different income levels.

As for the between-group components, high inequalities can be driven by group means far apart or by overlap (see discussion in Section 2.4). With the region-based scaling in Table 9, we see that some of the between-group inequalities are very large indeed. The largest two terms (89.4 between Africa and Northern America and 89.7 between Northern America and South/West/Central Asia) reflect high differences in mean income. Other differences, such as that between Latin America and East/Southeast Asia at 58.9, are more driven by overlaps between two regions with high internal inequalities.

3.4 Inequality since 1820

From the above discussion, we conclude that inequality between rich and poor countries, and in particular between Europe/North America and Asia, is the largest contributor to global inequality. Has this always been the case? Using data from Bourguignon and Morrisson (2002), who estimated global inequality for a set of years between 1820 and 1992, we can also look at historical inequality, using the same regions as used for the 2005 data. [10]

[Table 10 about here.]

[10] For data details, see the Appendix. As Bourguignon and Morrisson (2002) used aggregated country groups to estimate global inequality, we are not able to distinguish Oceania from Northern America. For this reason, this part of the analysis consists of six regions rather than seven. Moreover, for the same reason, all the unspecified Asian countries in BM's data (which does not include India or Indonesia) had to be grouped with "South/West/Central Asia".

Table 10 shows the population shares and relative mean income for a selection of years, based on a re-grouping of the data of Bourguignon and Morrisson. In terms of contribution to the Gini coefficient, we are mainly interested in the product of population share and mean income level. In terms of population, East/Southeast Asia is the largest region in all periods, though there is a substantial fall in population shares in the first 90 years, from 42% in 1820 to 32% in 1910. There is also a substantial fall in relative mean incomes as North America and Europe pulls ahead; in 1820, East/Southeast Asian mean income was at 52 per cent of that in Europe, while this proportion had decreased to 23 per cent by 1950. In the last half-century, however, there is substantial convergence between East/Southeast Asia and Europe and Northern America. For some of the other regions, such as Africa, there is no sign of mean income convergence at the regional level. In terms of population, North America experienced substantial growth in the early period, from 1% of world population in 1820 to 6% in 1910, while Africa's share of world population has grown from 6% in 1910 to 12% in 1992.

We can then examine how these differences in population and mean income translate to the development the components of global income inequality. The long-term evolution of the contributions to world Gini are given in Figure 2. Only terms that at any point in time contributed more than two Gini points are included in the figure. According to Bourguignon and Morrisson's numbers, global income inequality between individuals increased from a Gini coefficient of 49.7 in 1820 to 65.8 in 1992, with a nearly monotonous increase. However, the decomposition shows that several opposing trends underlie the smooth aggregate movement.

[Figure 2 about here.]

Income differences between Europeans and East/Southeast Asians constitute the largest contribution to global inequality today, and as shown in Figure 2, it has done so since 1820. The importance of this term has, however, declined over time, from more than one fourth of global inequality in 1850 to less than one eighth in 1992. This decrease reflects both Europe's declining global population share and Asian economic growth relative to Europe. The between-group term comparing Europe and South/West/Central Asia shows a similar trend; there is, however, less of a decline over time as income disparities between these regions have remained high.

The interaction between Northern America and the two Asian regions shows an almost linear increase between 1820 and 1950. In this period, income in North America grew considerably compared to the rest of the world. After 1950, there has, at least for some countries, been a reduction in income disparities, which reduces the contribution to the Gini coefficient.

Much of the increase in global inequality after 1950 comes from within-Asia inequality. Interestingly, this was also an important part of global inequality in 1820. Until 1950, both the two within-Asia components and the interaction between East/Southeast and West/South/Central Asia declined considerably, but after 1950 there has been a strong increase. This reflects both the time trend in the relative population of Asia and income differences between Asian countries.

Within-Europe inequality accounted for more than ten per cent of global inequality in 1910, compared to less than five per cent today. Catch-up has reduced within-Europe disparities at the same time as Europe's importance in the world - measured as the share of global population - has fallen strongly. Similarly, increasing population in Africa has increased the contribution of income difference between African individuals and between Africans and the rest of the world, though the total contribution of Africa remains as low as 11 Gini points in 1992 (of which 3 is Within-Africa inequality).

4 Discussion

4.1 Comparison to existing studies

The decomposition of global inequality presented here is novel in that it decomposes global inequality into a set of additive terms. As such, it cannot be directly compared to previous decompositions of global inequality, though there are some similarities.

The importance of the relationship between large, non-rich Asian countries and small, rich Western countries feature in several of the existing studies of global inequality. For example, Milanovic (2002) state that in 1993, the largest contributions to global inequality came from the very big countries, such as India and China, and the very rich countries. The role of the difference in mean income between India, China and the United States is also highlighted by Milanovic (2005, p. 88f).

There are two existing studies using a full decomposition of global inequality, both in the framework of Yitzhaki (1994) utilizing a term that categorizes the overlap between groups. Milanovic and Yitzhaki (2002) find that world regions, even when constructed as "economic-political groupings" by using shared historical background as a grouping criterion in addition to geography, does not classify the world as well as an "old-fashioned" partition into rich, middle income and poor countries. However, if one sticks to such a partition by said groupings, Asia is the largest contributor to world inequality both through large internal differences and a large overlap with the rest of the world distribution, while the contributions of Europe and North America are more modest. Liberati (2015) examines the time trend in global inequality between 1970 and 2009, and finds that while within-country inequality (defined as in Yitzhaki (1994), and hence not directly comparable to the within-terms used in the present paper) and overlapping of distributions between countries have become increasingly important during this time period, there have only been moderate changes in the global income Gini during this period.

As for the historical development of inequality, Bourguignon and Morrisson (2002) highlights the role of cross-region growth differences in the very long run, combined with Europe's ascension.[11] The discussion of the 1988-1993 time span in Milanovic (2002) sketches a remarkably similar development, with the exception that several countries in East Asia, as well as parts of urban China, now belong in the "rich" world.

Milanovic (2011) argues that global inequality in the early nineteenth century was mainly driven by inequality within countries, while twenty-first century inequality is driven by inequality between countries; we have gone from a class-divided world into a location-divided world. To some extent, this paper agrees with that idea; however, Figure 2 shows that the within-region differences have always been quantitatively important.

This article has established that a substantial portion of the global Gini coefficient stems from the differences between the high-mean-income regions of Europe and North America on one side and Asia on the other. If we compare the shape of the global income distribution to that of a country, we observe that the upper tail is very "thick"; this thickness, which is in this paper mainly allocated to comparisons including individuals in Europe and North America (though there are also countries in other regions with high incomes) is the main reason while the global Gini coefficient is higher than that typically seen for within-country distributions.

4.2 Concluding comments

This article has used a pairwise decomposition method for the Gini coefficient to show that the majority of the contribution to world inequality, both today and historically, comes from inequality within Asia and between Asia and "the West". Even though a large share of the world's extremely poor live in sub-Saharan Africa (around one third, according to Chen and Ravallion, 2010), Africa is not populous enough to affect the world's Gini coefficient by a large

[11] Bourguignon and Morrisson do not decompose the world Gini, rather relying on decompositions of the Theil index and the mean logarithmic deviation.

number.

While income distributions and economic systems vary significantly both among rich European, North American and Asian countries and among poor Asian countries, the inequality contributions between regions with large differences in mean income are remarkably similar. In 2005, there was still a rather low overlap between the distributions of poor and rich countries, meaning that differences in region-mean incomes drive a lot of global inequality. If the sustained growth that has taken place in Asian countries over the last ten years continues, this is likely to change, and the within-country distributions will have a larger impact on global inequality.

A Appendix

A.1 Data

Two existing data sets were used for computing the measures in Section 3.

The 2005 data is from Milanovic (2010). Income data in LCU (local currency units) were downloaded from http://econ.worldbank.org/projects/inequality; these were combined with ICP price conversion and inflation data from http://data.worldbank.org as well as ICP conversion rates for some more countries kindly supplied by Milanovic.

The historical data (BM henceforth) is from Bourguignon and Morrisson (2002), downloaded from http://www.delta.ens.fr/XIX.

A.1.1 Using quantile data

Both data sets use quantile data (mostly 20 groups for Milanovic and always 11 groups for BM), meaning that the complete income distribution is collapsed and concentrated at a finite number of income points. Milanovic (2010) has some discussion of the impreciseness of this method and shows that for a selection of countries where better data is available, the simplification does not affect country Ginis by more than one per cent. This is potentially a larger problem with the BM data, which only has eleven income groups. However, to stay consistent with the results in Bourguignon and Morrisson (2002), no smoothing of the distributions were used.

A.1.2 Interpolating the Milanovic data

The Milanovic data covers most of the world population, and while the omissions are more prevalent among poor countries, it does not contribute much to the aggregate Gini calculation. When decomposing by continent or subregion, however, it is important to have unbiased estimates of population and income sizes for these regions; otherwise Africa, in particular, would be weighted down. For this reason, the following interpolation methods were used to extend the dataset to cover the entire world. The countries can be grouped according to the following criteria:

Group 0: no interpolation. Income distribution available from the Milanovic data set, ICP and other PPP data available. Total population in this group is 5836 million; 118 countries (of which China, separated into urban and rural, counts as two countries).

Group 1: unknown income distribution in 2005, but available for 1998 or 2002. The incomes are scaled by the difference in PPP GDP per capita (according to the World Bank) between the data year and 2005. Total population in the 16 countries interpolated from 2002 is 218 million; from 1998, a total of 7 countries with a population of 35 million was used.

Group 2: unknown income distribution. In this case the income distribution was taken from a neighboring country, usually one bordering the country in question, in the same subregion and with comparable GDP levels. The levels of income were then scaled by the difference in PPP GDP per capita between the interpolated and the original country. Total population in Group 2 is 293 million; 48 countries. This includes some countries with missing GDP data in the World Bank data; in that case, differences between the country in question and the US were taken from the Penn World Tables or, in three cases, the CIA World Factbook. For West Sahara no GDP estimate was found; GDP per capita of West Sahara was assumed to be equal to that in Mauritania.

In addition, some very small countries with population were dropped from the sample altogether, mostly very minor Caribbean and Pacific islands with five-digit populations.

A.2 Inequality decomposition 1820-1992, all terms

See Table 11.

[Table 11 about here.]

References

Anand, S. and P. Segal, "What Do We Know about Global Income Inequality?," *Journal of Economic Literature*, 46, 57–94, 2008.

Atkinson, A. B. and A. Brandolini, "Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries as a Case Study," *Journal of Economic Literature*, 39, 771–799, 2001.

Bourguignon, F. and C. Morrisson, "Inequality among World Citizens: 1820-1992," *American Economic Review*, 92, 727–744, 2002.

Chen, S. and M. Ravallion, "The Developing World is Poorer than We Thought, But No Less Successful in the Fight Against Poverty," *The Quarterly Journal of Economics*, 125, 1577–1625, 2010.

Chotikapanich, D., R. Valenzuela, and D. S. P. Rao, "Global and regional inequality in the distribution of income: Estimation with limited and incomplete data," *Empirical Economics*, 22, 533–546, 1997.

Dikhanov, Y. and M. Ward, "Evolution of the Global Distribution of Income, 1979-99," *Paper prepared for the 53rd Session of the International Statistical Institute, Seoul, Republic of Korea, August 22-29, 2001*, 2001.

Dowrick, S. and M. Akmal, "Contradictory trends in global income inequality: a tale of two biases," *Review of Income and Wealth*, 51, 201–229, 2005.

Ebert, U., "The decomposition of inequality reconsidered: Weakly decomposable measures," *Mathematical Social Sciences*, 60, 94–103, 2010.

Lambert, P. J. and J. R. Aronson, "Inequality Decomposition Analysis and the Gini Coefficient Revisited," *Economic Journal*, 103, 1221–1227, 1993.

Liberati, P., "The world distribution of income and its inequality, 1970-2009," *Review of Income and Wealth*, 61, 248–273, 2015.

Milanovic, B., "True world income distribution, 1988 and 1993: First calculation based on household surveys alone," *Economic Journal*, 112, 51–92, 2002.

_ , *Worlds Apart: Measuring International and Global Inequality*, Princeton University Press, 2005.

_ , "Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates," *Journal of Economic Inequality*, 10, 1–18, 2010.

— , “A short history of global inequality: The past two centuries,” *Explorations in Economic History*, 48, 494–506, 2011.

— and S. Yitzhaki, “Decomposing World Income Distribution: Does The World Have A Middle Class?,” *Review of Income and Wealth*, 48, 155–178, 2002.

Modalsli, J., “Inequality in the very long run: inferring inequality from data on social groups,” *Journal of Economic Inequality*, 13, 225–247, 2015.

Mookherjee, D. and A. Shorrocks, “A Decomposition Analysis of the Trend in UK Income Inequality,” *The Economic Journal*, 92, 1982.

Pyatt, G., “On the Interpretation and Disaggregation of Gini Coefficients,” *The Economic Journal*, 86, 1976

Shorrocks, A. F., “Inequality Decomposition by Population Subgroups,” *Econometrica*, 52, 1369–1385, 1984.

Yitzhaki, S., “Economic distance and overlapping of distributions,” *Journal of Econometrics*, 61, 147–159, 1994.

— and R. I. Lerman, “Income stratification and income inequality,” *Review of Income and Wealth*, 37, 313–329, 1991.

Figure 1: Global Gini decomposition: Map. Circles: Within-region inequality; rectangles: between-region inequality. Sums to 69.7 (global Gini coefficient)

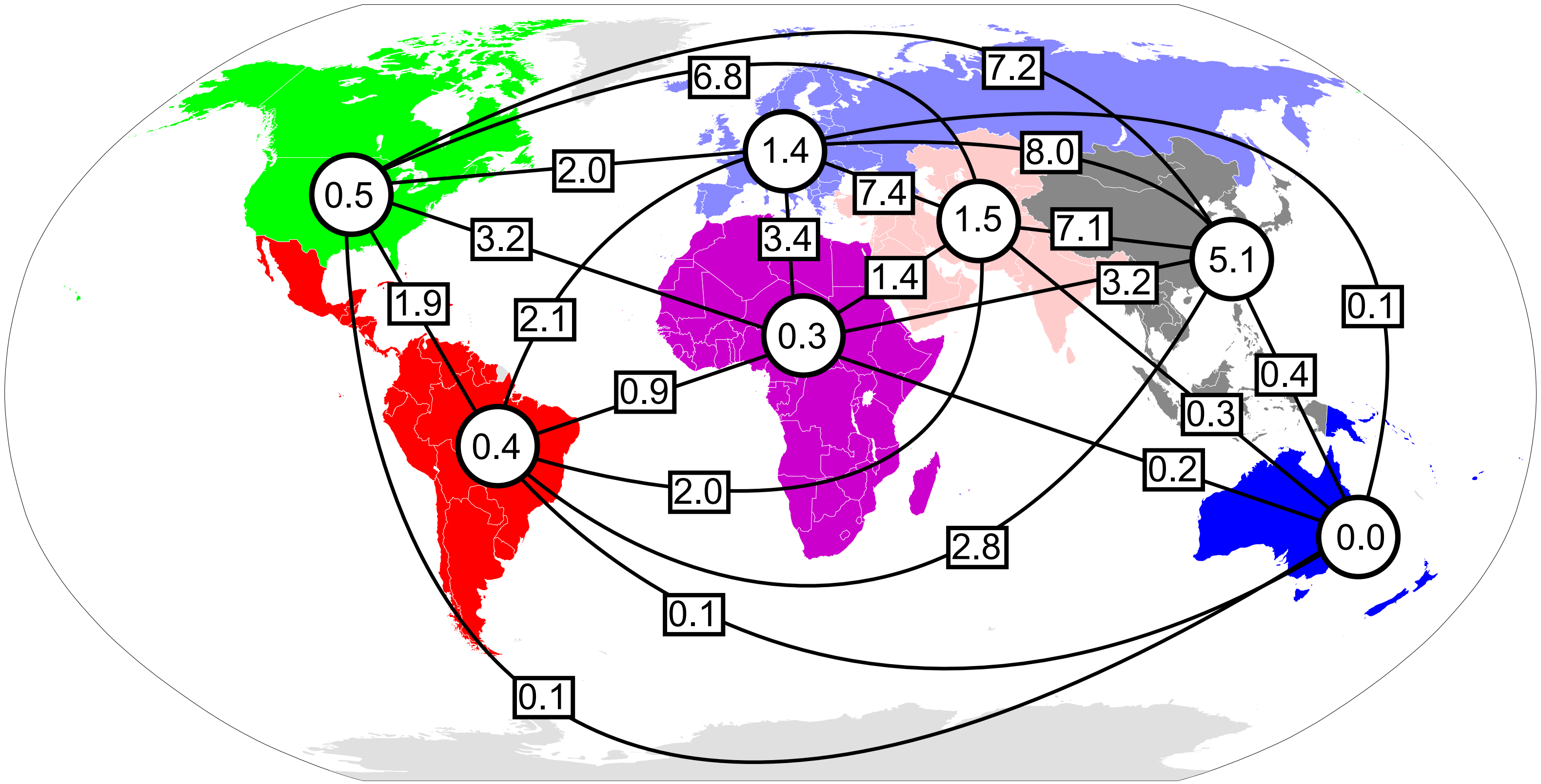

Figure 2: Historical inequality

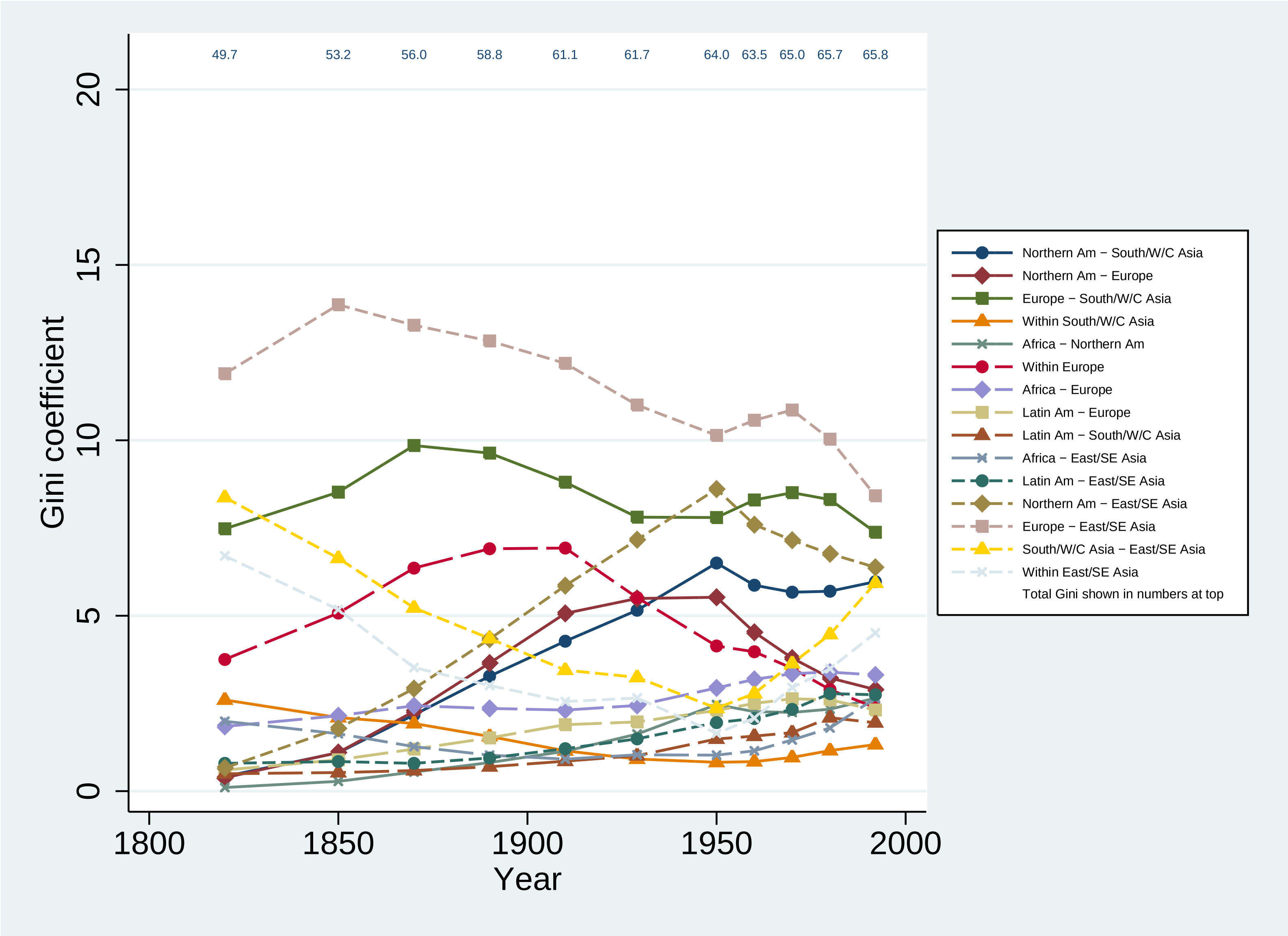

Table 1: Decomposition example: Difference tabulation

		$r=1$			$r=2$		$r=3$	
		2	5	8	5	11	4	7
	2	0						
$q=1$	5	3	0					
	8	6	3	0				
$q=2$	5	3	0	3	0			
	11	9	6	3	6	0		
	4	2	1	4	1	7	0	
$q=3$	7	5	2	1	2	4	3	0

Table 2: Decomposition example: Sum of group differences

	1	2	3
1	12		
2	24	6	
3	15	14	3

Table 3: Decomposition example: Contribution of each group

	1	2	3
1	4.1		
2	8.2	2.0	
3	5.1	4.8	1.0

	1	2	3
1	4.1		
2	8.2	2.0	
3	5.1	4.8	1.0

	1	2	3
1	4.1		
2	8.2	2.0	
3	5.1	4.8	1.0

Table 4: Decomposition example: Mean-between and residual components

	1	2	3
1	$G^r = 4.1$		
2	$G^m = 6.1$ $G^r = 2.0$	$G^r = 2.0$	
3	$G^m = 1.0$ $G^r = 4.1$	$G^m = 3.4$ $G^r = 1.4$	$G^r = 1.0$

Table 5: Contributions to global inequality: largest terms in decomposition by country

Country pair	Share of global Gini	G_{qr}	Betw. G_{qr}^m	Resid. G_{qr}^r	Pop. w. $p_q p_r$	Inc. w. $\left(\frac{\mu_q}{\mu}+\frac{\mu_r}{\mu}\right)$	Group-scaled Gini G_{qr}^w
China and USA	6.0%	4.2	4.2	0.0	0.009	5.443	81.6
India and USA	5.5%	3.8	3.8	0.0	0.008	5.063	94.6
China and Japan	2.1%	1.4	1.4	0.0	0.004	4.563	77.5
China and India	2.0%	1.4	1.3	0.1	0.035	0.651	62.3
India and Japan	1.9%	1.3	1.3	0.0	0.003	4.183	93.5
China and Germany	1.4%	1.0	1.0	0.0	0.003	4.630	77.9
China (within)	1.3%	0.9	-	0.9	0.042	1.031	42.0
Germany and India	1.3%	0.9	0.9	0.0	0.002	4.251	93.6
Indonesia and USA	1.1%	0.8	0.8	0.0	0.002	5.163	90.9
China and UK	1.0%	0.7	0.7	0.0	0.002	4.755	78.7
UK and India	1.0%	0.7	0.7	0.0	0.002	4.375	93.8
China and France	0.9%	0.6	0.6	0.0	0.002	4.329	76.5
France and India	0.9%	0.6	0.6	0.0	0.002	3.949	93.1
Brazil and USA	0.8%	0.6	0.5	0.0	0.001	5.874	72.6
Bangladesh and USA	0.8%	0.5	0.5	0.0	0.001	4.961	98.6
Pakistan and USA	0.8%	0.5	0.5	0.0	0.001	5.147	91.5
Nigeria and USA	0.7%	0.5	0.5	0.0	0.001	5.040	95.5
Brazil and China	0.7%	0.5	0.3	0.2	0.006	1.462	54.4
China and Italy	0.6%	0.4	0.4	0.0	0.002	3.315	70.3
USA (within)	0.6%	0.4	-	0.4	0.002	9.854	40.3

Table 6: Regions used in the analysis

	Population (millions)	Population (relative)	Income per capita (world=100)
Africa	858	13%	29
Latin America and the Caribbean	555	9%	86
Northern America	329	5%	486
Europe (including Russia)	705	11%	254
Asia: South, West, Central	1817	28%	29
Asia: East, South East	2087	33%	79
Oceania	33	1%	247

Table 7: Global Gini decomposition

	Africa	Latin America and Caribbean	Northern America	Europe	Asia: South, West, Central	Asia: East, Southeast	Oceania
Africa	0.3						
Latin America and Caribbean	0.9	0.4					
Northern America	3.2	1.9	0.5				
Europe	3.4	2.1	2.0	1.4			
Asia: South, West, Central	1.4	2.0	6.8	7.4	1.5		
Asia: East, Southeast	3.2	2.8	7.2	8.0	7.1	5.1	
Oceania	0.2	0.1	0.1	0.1	0.3	0.4	0.0

Total Gini: 69.7

Table 8: Global Gini decomposition: means and residuals. Adding M and R components within each cell gives Table 7.

	Africa	Latin America and Caribbean	Northern America	Europe	Asia: South, West, Central	Asia: East, Southeast	Oceania
Africa	M: 0.0 R: 0.3						
Latin America and Caribbean	M: 0.7 R: 0.2	M: 0.0 R: 0.4					
Northern America	M: 3.2 R: 0.0	M: 1.8 R: 0.1	M: 0.0 R: 0.5				
Europe	M: 3.3 R: 0.1	M: 1.6 R: 0.4	M: 1.3 R: 0.7	M: 0.0 R: 1.4			
Asia: South, West, Central	M: 0.0 R: 1.3	M: 1.4 R: 0.6	M: 6.7 R: 0.1	M: 7.1 R: 0.4	M: 0.0 R: 1.5		
Asia: East, Southeast	M: 2.2 R: 1.0	M: 0.2 R: 2.6	M: 6.9 R: 0.4	M: 6.3 R: 1.6	M: 4.6 R: 2.4	M: 0.0 R: 5.1	
Oceania	M: 0.1 R: 0.0	M: 0.1 R: 0.0	M: 0.1 R: 0.0	M: 0.0 R: 0.1	M: 0.3 R: 0.0	M: 0.3 R: 0.1	M: 0.0 R: 0.0

Table 9: Inequality scaled by group means and population sizes

	Africa	Latin America and Caribbean	Northern America	Europe	Asia: South, West, Central	Asia: East, Southeast	Oceania
Africa	57.7						
Latin America and Caribbean	67.4	56.6					
Northern America	89.4	73.7	39.8				
Europe	82.1	62.9	48.2	44.8			
Asia: South, West, Central	60.9	70.3	89.7	83.2	63.4		
Asia: East, Southeast	67.0	58.9	76.0	66.0	69.8	60.6	
Oceania	83.2	65.3	49.1	46.7	84.1	68.1	48.0

Table 10: Historical population shares and mean income levels. (Oceania is included with South/West/Central Asia).

	Population share					Mean income (world=100)				
Region	1820	1870	1910	1950	1992	1820	1870	1910	1950	1992
Africa	7%	7%	6%	9%	12%	76	62	49	42	29
Latin America and Caribbean	2%	3%	4%	6%	8%	106	83	95	116	97
Northern America	1%	4%	6%	7%	6%	188	274	337	432	420
Europe	22%	26%	28%	24%	16%	157	174	159	166	195
Asia: South, West, Central	27%	26%	23%	23%	27%	82	64	47	33	36
Asia: East, Southeast	42%	35%	32%	31%	31%	83	63	52	38	79

Table 11: Historical inequality: Gini contributions from the five regions, 1820-1992 (as illustrated in Figure 2)

	Within Africa	Africa - Latin Am	Africa - Northern Am	Africa - Europe	Africa - South/W/C Asia	Africa - East/SE Asia	Within Latin Am	Latin Am - Northern Am	Latin Am - Europe	Latin Am - South/W/C Asia	Latin Am - East/SE Asia	Within Northern Am	Northern Am - Europe	Northern Am - South/W/C Asia	Northern Am - East/SE Asia	Within Europe	Europe - South/W/C Asia	Europe - East/SE Asia	Within South/W/C Asia	South/W/C Asia - East/SE Asia	Within East/SE Asia
1820	0.1	0.1	0.1	1.8	1.2	2.0	0.0	0.0	0.6	0.5	0.8	0.0	0.4	0.4	0.7	3.8	7.5	11.9	2.6	8.4	6.7
1850	0.1	0.1	0.3	2.2	1.0	1.6	0.0	0.1	0.9	0.5	0.8	0.1	1.1	1.1	1.8	5.1	8.5	13.9	2.1	6.6	5.2
1870	0.1	0.1	0.5	2.4	0.9	1.3	0.0	0.3	1.2	0.6	0.8	0.2	2.3	2.2	2.9	6.4	9.9	13.3	1.9	5.2	3.5
1890	0.1	0.2	0.8	2.4	0.7	1.0	0.1	0.5	1.5	0.7	0.9	0.4	3.7	3.3	4.3	6.9	9.6	12.8	1.6	4.4	3.0
1910	0.1	0.2	1.1	2.3	0.6	0.9	0.1	0.8	1.9	0.9	1.2	0.6	5.1	4.3	5.9	6.9	8.8	12.2	1.1	3.5	2.6
1929	0.1	0.3	1.6	2.4	0.6	1.0	0.2	1.1	2.0	1.0	1.5	0.8	5.5	5.2	7.2	5.5	7.8	11.0	0.9	3.2	2.7
1950	0.2	0.6	2.5	2.9	0.7	1.0	0.3	1.7	2.3	1.5	2.0	0.9	5.5	6.5	8.6	4.1	7.8	10.1	0.8	2.4	1.6
1960	0.2	0.6	2.3	3.2	0.7	1.1	0.3	1.6	2.5	1.6	2.1	0.8	4.5	5.9	7.6	4.0	8.3	10.6	0.8	2.8	2.1
1970	0.2	0.7	2.2	3.4	0.8	1.5	0.3	1.6	2.6	1.7	2.3	0.7	3.8	5.7	7.2	3.5	8.5	10.9	1.0	3.6	3.0
1980	0.2	0.8	2.3	3.4	0.9	1.8	0.4	1.7	2.6	2.1	2.8	0.6	3.2	5.7	6.8	2.9	8.3	10.0	1.2	4.5	3.5
1992	0.2	0.9	2.7	3.3	1.1	2.6	0.4	1.7	2.3	2.0	2.7	0.6	2.9	6.0	6.4	2.4	7.4	8.4	1.3	5.9	4.5