Inequality in the very long run: inferring inequality from data on social groups

Jørgen Modalsli

The final authenticated version is available at:

Modalsli, J. (2015). Inequality in the very long run: inferring inequality from data on social groups. *The Journal of Economic Inequality,13(2)*, 225-247. doi: https://doi.org/10.1007/s10888-014-9279-6

Inequality in the very long run: inferring inequality from data on social groups

Jørgen Modalsli, Statistics Norway*

June 27, 2014

Abstract

This paper presents a new method for calculating Gini coefficients from tabulations of the mean income of social classes. Income distribution data from before the Industrial Revolution usually come in the form of such tabulations, called social tables. Inequality indices generated from social tables are frequently calculated without adjusting for within-group income dispersion, leading to a systematic downward bias in the reporting of pre-industrial inequality.

The correction method presented in this paper is applied to an existing collection of twenty-five social tables, from Rome in AD 1 to India in 1947. The corrections, using a variety of assumptions on within-group dispersion, lead to substantial increases in the Gini coefficients.

Keywords: Pre-industrial inequality, social tables, Kuznets curve, history

JEL codes: D31, N30, O11, C65

---

*E-mail: mod@ssb.no. This paper is part of the research activities at the centre of Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo. ESOP is supported by the Research Council of Norway. I am grateful to Rolf Aaberge, Gernot Doppelhofer, Livio Di Matteo, Halvor Mehlum, Branko Milanovic, Kalle Moene, Erik Sørensen, the Editor and two anonymous referees for comments and suggestions.

1 Introduction

Not much is known about inequality in the very long run. The lack of data has been addressed by Milanovic *et al.* (2011), who collect a large set of social tables. The social tables give data on the size and average income of social classes in many pre-industrial societies, with the catch that the income distribution within each class is unknown. This paper shows that common approaches to dealing with this problem do not take sufficient account of within-group inequality, which might lead to downward biased Gini coefficient estimates. For this reason, a new approach is developed in Section 2. In Section 3, this approach is applied to the data of Milanovic *et al.*, leading to a large upward revision of the estimates of inequality.

1.1 Inequality in the very long run

The seminal contribution on the long-run evolution of inequality is Kuznets (1955). Using a few observations from the United States, England and Germany, Kuznets argues that inequality goes up with the industrial revolution and then decreases with modernization. While Kuznets treats the Industrial Revolution as a rather specific process (he dates the possible "widening phase" in England as going from 1780 to 1850, and postulates even shorter periods for the other countries), more recent views on industrialization stress the changes as being more gradual.

Kuznets based his conclusions on a very small data set. Over the years, more data points have become available. For example, Van Zanden (1995) reports Gini coefficients for many European cities before from the 1500s onward, Lindert (2000) analyze inequality in Britain and the United States after 1700, and Hoffman *et al.* (2002) report Gini coefficients for several European countries. An early meta-study is that by Bourguignon & Morrisson (2002), who combine inequality data for various countries to construct an estimate of the world income inequality from 1820 onwards.

The most comprehensive analysis of pre-industrial inequality so far is given by Milanovic *et al.* (2011). The authors collect a comprehensive set of social tables - listing social groups, their sizes and incomes for 24 country-time points. An example of a social table is given in Table 1. It lists the social classes in Byzantium,

Social group	Share of pop.	Per capita income (nomisma per year)	Income in terms of per capita mean
Tenants	0.37	3.5	0.56
Urban “marginals”	0.02	3.51	0.56
Farmers	0.52	3.8	0.61
Workers	0.03	6	0.97
Army	0.01	6.5	1.05
Traders, skilled craftsmen	0.035	18	2.90
Large landowners	0.01	25	4.02
Nobility	0.005	350	56.31

Table 1: Example of social table: Byzantium, ca year 1000. Source: Milanovic *et al.* (2007), based on Milanovic (2006)

ca year 1000. The data set used in this paper consists of 24 such social tables, with a varying number of groups and class definitions.[1] Though far from being a balanced panel (only a few countries have observations for more than one period), this is the first comprehensive cross-region data series on pre-industrial inequality, as opposed to the more country- or region-specific discussions of the other studies.

1.2 Interpolating inequality: Limitations of existing approaches

Common for all elaborations on pre-industrial inequality is the need for some type of interpolation. Often a combination of techniques is used, as the data available can be of many types. For example, Lindert (2000) uses a combination of social tables, factor prices, wage data, and land holdings, as well as more detailed data on wealth and income for the richer parts of the population. In most cases, information on the distribution among the poor is particularly hard to find.

For the social tables collected by Milanovic *et al.* (2011), we have the advantage

[1]Milanovic *et al.* have a total of 28 observations. For two of these (Holland 1561 and Japan 1886) they do not appear to have access to the underlying data. For another two (Tuscany 1427 and Bihar 1807) the data is not available in a format based on social groups. For the remaining 24 observations, based on a wide range of studies described in their paper, I thank Branko Milanovic for supplying the dataset; most of the observations are also available online at `http://gpih.ucdavis.edu/`. The working paper version of their paper (Milanovic *et al.*, 2007) has a fuller exposition of the data and methodology.

of a comprehensive table for the entire population.[2] For each social class, we have an estimate of mean income of the group, as well as the relative size of the group. The distribution within each group, however, is not known. For this reason, analyzing inequality using social tables data requires additional assumptions on the characteristics of the social groups.

A natural starting point is to consider a distribution where the entire group is concentrated at its mean income. Taking the "farmers" in Table 1 as an example, this would mean that all farmers had an income of 3.8 nomisma per year. This assumption makes it easy to calculate an inequality measure such as the Gini coefficient. Milanovic *et al.* (2011) describe this as the lower bound of the Gini coefficient, and denote it as "Gini1". In the following, this will be referred to as a "point distribution", as the population is concentrated at a finite number of points.[3]

Going one step further, we can think of a distribution where all the members of group i are poorer than all members of group $i+1$; in the terms of Table 1, all "tenants" are poorer than the poorest farmer. This will be referred to as a population being *perfectly sorted* by groups; in other words, there is no overlap between the population ranges. The highest inequality consistent with this assumption is found for a distribution with half of the individuals in each group having income at the lower border, and the other half at the upper border. For group borders at midpoints between group means, Milanovic *et al.* (2011) denote this as "Gini2", but alternatively we could also conceive a situation where we set the group borders so as to *maximize* the inequality consistent with the assumption of perfect sorting.

For most social table distributions, the assumption of perfect sorting greatly limits the possible Gini coefficients. An illustration of this is shown in Figure 1, which shows the Lorenz curve for a population of four groups. The Lorenz curve plots cumulative population against cumulative income, and the area between the Lorenz curve and the 45-degree line is equal to the Gini coefficient of the population. When groups are perfectly sorted, the points $(0, 0)$, (P_1, Z_1), ... are

[2]There is of course substantial uncertainty inherent in compiling the tables. This goes for any pre-industrial data series, including wage and other price series, and will not be discussed further here.

[3]Analytical expressions will be detailed below; the "point distribution" Gini is equal to the between-group Gini, given in Equation (7).

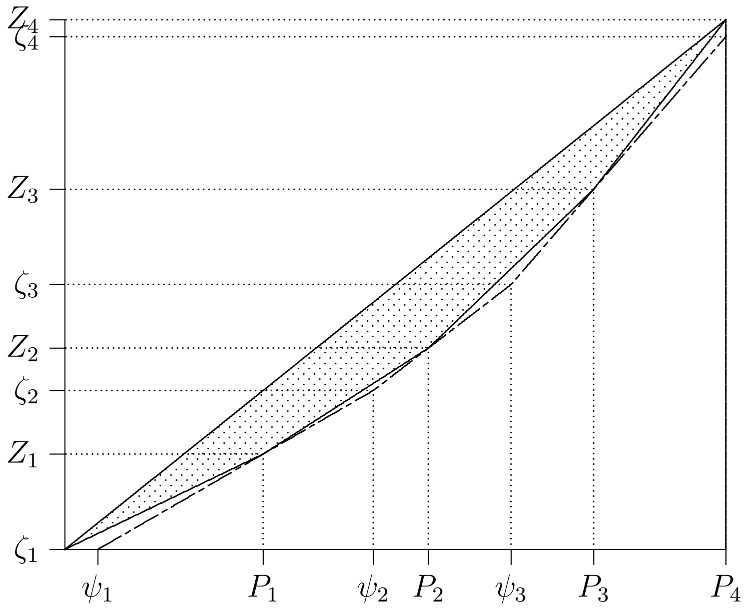


Figure 1: Lorenz curve and Gini coefficients for two restrictive assumptions

known; (P_i, Z_i) refers to the cumulative population and income of all groups up to group i. If there is no dispersion within groups, the Lorenz curve is given by the solid line, and the minimum Gini is the shaded area in the figure.

Now consider a set of within-group dispersions that preserves the perfect ordering of incomes by groups. The points (P_i, Z_i) still have to be on the Lorenz curve. Moreover, by the definition of the Lorenz curve, it must always be weakly convex — the Lorenz curve plots population sorted by income, and the slope of the curve corresponds to the income of an individual at that point. It follows that the most outward-lying Lorenz curve is a series of straight lines going through the points (P_i, Z_i) with kinks somewhere between these points; an example of such a line is the dotted line in Figure 1. Correspondingly, the Gini coefficient can only go up by the area between the solid and dotted line.[4]

The max-inequality Lorenz reflects a distribution where the population of a group is concentrated at the two extremes of the income groups' range; the richest individuals in group i have the same income as the poorest in group $i + 1$. The

[4]A related analytical proof for the case when group interval borders are given is found in Gastwirth (1972).

position of these income and population points, denoted (ψ_i, ζ_i) in the figure, that gives the highest possible Gini is in general not easy to find in closed form. However, as is evident from the figure, for most distributions the scope of increasing the area between the solid and dotted lines is very limited, and becomes more so as the number of groups goes up.

For a few "pre-industrial" societies, we do have information on inequality both within and across groups. This does allow for some examination of whether the restrictions described here are empirically plausible.

1.3 Overlaps between groups in pre-industrial societies

Of the 28 income distributions used by Milanovic *et al.*, two allow for more detailed analysis of within-group distributions.

The estimate for **Tuscany, 1427** uses data from the full-count Catasto (tax census). While the income estimates used by MLW appends wage data taken from other sources (without within-group information), the Catasto itself has wealth data and makes possible a full-count estimation of aggregate and decomposed wealth Gini coefficients.

The second source is the expenditure survey of **Bihar, 1807**. While there is no combined table with both social class/occupation and expenditure, expenditures are reported separately for rural and urban locations.

A third source, not used by Milanovic *et al.*, is a report containing income distributions for **Norway, 1868**. For a set of 26 occupational groups, the number of adult males earning above a threshold level is given, separated into five income groups. From this data we can construct aggregate and decomposed income Gini coefficients, contingent on earning above the threshold level. While the data only covers the upper third of the adult-male income distribution, it still gives valuable influence on the overlaps between groups in this income range.

The commonly used decomposition of the Gini coefficient, used, for example, by Lambert & Aronson (1993), divides total inequality into three components. Between-group inequality, G_B, follows directly from group means and is the inequality that the population would have if there was no inequality within groups. Within-group inequality, G_W, is a weighted sum of the Gini coefficient each group would have if it was a separate population. The remainding inequality, which is zero if there is no overlap between groups, is often referred to as "residual inequality" and will be denoted G_R. It is worth noting that the restriction of "no overlap" not only affects G_R, but also puts bounds on the within-group inequality.

Country	Unit	# groups	G	G_B	G_W	G_R
Tuscany, 1427	Wealth	97 occupations	75.2	46.5	19.4	9.3
Bihar, 1807	Expenditure	2 sectors	35.3	2.1	29.2	4.1
Norway, 1868	Income (upper 1/3)	26 occupations	29.2	15.2	5.9	8.1

Table 2: Pre-industrial societies with within-group data

For the three pre-industrial societies for which we have data, the three components of the Gini coefficient can be calculated separately, as shown in Table 2. It is clear that between-group inequality only accounts for a small part of inequality in these three societies. The extreme example is Bihar, where two large groups have means that are very close, but for the two other samples there is also substantial within-group inequality.

Even though the overlap term (G_R) is moderate the restriction of "no overlap" would lead to Gini coefficients much lower than the actual distributions. To see this, consider the methods of Section 1.2 applied to the three data sets, as shown in Table 3.

Country	Gini with point distribution (G_B)	Max Gini with no overlap	"True" Gini
Tuscany, 1427	46.5	52.9	75.2
Bihar, 1807	2.1	19.6	35.3
Norway, 1868	15.2	15.4	29.2

Table 3: Inequality with and without overlap

For each country, everyone were given their group mean income and inequality was calculated. This is the first column. The second column gives the Gini coefficient with the maximum dispersion consistent with "no overlap". The final column gives the Gini calculated from micro data. It is evident from the table that the limitation of "no overlap" is severe; in all cases, the difference between the group-calculated Ginis and the true Ginis are more than 10. This highlights the importance of relaxing the no-overlap restriction when calculating inequality from group data.

The limitation of assuming perfectly sorted groups, if this does not correspond to known characteristics of the underlying population, is the main motivation for imposing within-group distributions that have overlaps between the income ranges of groups. This will be the topic of the next section.

2 Social tables and log-normal group distributions

2.1 The distribution of income within groups

To put some structure on the within-group dispersion of income, it will be assumed for the remainder of this paper that income within each social class is log-normally distributed. The log-normal distribution is commonly used to model income inequality. For a stochastic process with a given population, where relative changes in incomes are random, the central limit theorem yields a log-normal distribution for this population (see, for instance, Crow & Shimizu (1987, chap. 1), citing Gibrat (1930, 1931)). If group incomes are log-normally distributed, the corresponding theoretical justification is that while the conventional stochastic processes operate within groups, there is no mobility between groups. The different means would be explained by a variety of different initial conditions "outside the model", unequal land distributions, historical conquests, discrimination or institutionalized privilages. While somewhat stylized, this is a reasonable and easily understood assumption, in particular on historical data.[5]

With log-normal distributions within groups, the aggregate distribution will not itself be log-normal. Rather, it captures the salient features of a presumably stratified society; the distribution shape will reflect the group data and its smoothness will depend on within-group dispersion. The log-normal distribution

[5]The pre-industrial distributions discussed in the previous section have some "bracketed" data within each group, making formal tests of distributional shapes difficult without further assumptions. However, some evidence points toward groupwise lognormality in these cases. See the Online Appendix for details.

has mass along the entire positive income range; correspondingly, there will be overlap between groups and the Lorenz curve will pass to the right of the points (P_i, Z_i) in Figure 1.

The log-normal distribution is most conveniently expressed in terms of μ, the mean of log income, and σ, the standard deviation of log income.Denoting the mean income of a group as y_i and the standard deviation of the income as s_i, the expressions for these parameters are

$$\mu_i = \log(y_i) - \frac{1}{2}\log\left(1 + \left(\frac{s_i}{y_i}\right)^2\right) = \log(y_i) - \frac{\sigma_i^2}{2} \tag{1}$$

$$\sigma_i^2 = \log\left(1 + \left(\frac{s_i}{y_i}\right)^2\right) \tag{2}$$

The cumulative distribution function (cdf) is

$$F^L(x; \mu, \sigma) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right) \tag{3}$$

where $\Phi(\cdot)$ is the standard cumulative normal distribution,
$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t$.

Denoting the relative population size of each group (social class) by p_i and the total number of groups by N, it follows that the cumulative income distribution function of the population is defined by

$$F(x) = \sum_{i=1}^{N} \left[p_i F^L(x; \mu_i, \sigma_i)\right] \tag{4}$$

where μ_i and σ_i are defined by (1) and (2).

2.2 Calculating Gini coefficients from group data

As demonstrated by Aitchison & Brown (1957), the Gini coefficient for the log-normal distribution (3) is given by $G^L = 2\Phi(\sigma/\sqrt{2}) - 1$. Using the procedure given in the Appendix, we can derive the Gini coefficient of the distribution F defined by (4). This gives a closed-form expression for the Gini coefficient that incorporates overlaps between groups.

Proposition 1 *Let a population with mean income $\bar{y}$ be divided into N groups where each group i has population share p_i and a log-normal income distribution with parameters (μ_i, σ_i^2), $i = 1, 2, ...N$. Then the Gini coefficient is given by*

$$G = \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j \frac{y_i}{\bar{y}} \left(2\Phi \left(\frac{\mu_i - \mu_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) - 1 \right) \tag{5}$$

Proof: *See Appendix.*[6]

This expression has N^2 terms; two for each combination of i and j. Each of the terms considers a separate part of the Lorenz square;[7] group i's share of income $p_i y_i / \bar{y}$ (on the vertical axis) is multiplied with group j's share of population p_j (on the horizontal axis). If there was no overlap, these parts would be separate rectangles and constitute a grid; however, in this case, the areas should be considered as density functions over the entire square. Each of these areas are weighted by a number between -1 and 1, depending on the corresponding values of μ and σ for the two groups. The sum of these weighted squares is a measure of the distance between all individuals; the Gini coefficient.

As the expression (5) has many more terms than the number of groups, and some of the terms are negative, it is not straightforward to interpret the effect of different parameters on the resulting Gini coefficient. For this reason, it is more convenient to work with a re-formulated expression. First, replace the parameter

[6] The relationship between group mean income y_i and (μ_i, σ_i^2) is given in Equations (1)-(2). Note that $\bar{y} = \sum_{i=1}^{N} p_i y_i$. To the knowledge of this author, the result in Equation (5) is not previously published. After the first working paper edition of this paper, Young (2011) has independently derived a similar expression, in the context of modern (national and global) income inequality.

[7] The term "Lorenz square" refers to the square on which the Lorenz curve is plotted; the horizontal axis represent aggregate population, sorted from poorest to richest, while the vertical axis represent cumulative aggregate income.

μ with the group means, using (1).[8] Second, add each ij term (where $i < j$) to the corresponding ji term to get the preferred expression for the Gini coefficient

$$G = \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} p_i p_j \left(\frac{y_j}{\bar{y}} \left[2\Phi\left(\frac{\log\left(\frac{y_j}{y_i}\right)}{\sqrt{\sigma_i^2+\sigma_j^2}} + \frac{\sqrt{\sigma_i^2+\sigma_j^2}}{2} \right) - 1 \right] - \frac{y_i}{\bar{y}} \left[2\Phi\left(\frac{\log\left(\frac{y_j}{y_i}\right)}{\sqrt{\sigma_i^2+\sigma_j^2}} - \frac{\sqrt{\sigma_i^2+\sigma_j^2}}{2} \right) - 1 \right] \right)}_{\text{Across-group inequality } (G_A = G_B + G_R)} + \underbrace{\sum_{i=1}^{N} p_i^2 \frac{y_i}{\bar{y}} \left[2\Phi\left(\frac{\sigma_i}{\sqrt{2}}\right) - 1 \right]}_{\text{Within-group inequality } (G_W)} \tag{6}$$

which is decomposed into an across-group inequality term (henceforth defined as $G_A = G_B + G_R$) and a within-group inequality term.[9]

The first term of (6) is the sum of inequality across groups; all pairwise comparisons between individuals in group i and individuals in group j. We can contrast this to the Gini coefficient for no within-group dispersion, which is the population-weighted sum of all pairwise differences between the groups

[8] One could also substitute in s for σ, but this does not add clarity; as the Gini coefficient is a relative measure, the standard deviation only enters scaled, as s/y, and this can just as well be summarized in the σ measure.

The Gini coefficient expressed only in means and standard deviations is

$$G = \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \frac{y_i}{\bar{y}} \left(2\Phi\left(\frac{\log\left(\frac{y_i}{y_j}\right)}{\sqrt{\log\left[\left(1+\frac{s_i^2}{y_i^2}\right)\left(1+\frac{s_j^2}{y_j^2}\right)\right]}} + \frac{\sqrt{\log\left[\left(1+\frac{s_i^2}{y_i^2}\right)\left(1+\frac{s_j^2}{y_j^2}\right)\right]}}{2} \right) - 1 \right)$$

[9] G_B, G_R and G_W were defined in Section 1.3. The decomposition into G_A and G_W is discussed by Ebert (2010), who treats G_A as the "between" component. The analysis here is also related to Yitzhaki & Lerman (1991), who study the relationship between stratification and inequality. The aggregate group data can be construed as giving stratification but not inequality, and the Gini coefficients presented here measure stratification-induced inequality differences between populations.

$$G_0 = \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} p_i p_j \left(\frac{y_j}{\bar{y}} - \frac{y_i}{\bar{y}}\right)}_{\text{Between-group inequality } (G_B)} \tag{7}$$

and see that the expressions are closely related. G_A differs from G_B in that the group means are modified by a number between -1 and 1; the evaluation of the $2\Phi(\cdot) - 1$ function.

The values for y and p in a given population are known from the social tables. The dispersion, however, is not. It is therefore of interest to know how the inequality of a population changes when dispersion changes - how G changes with s_i, or σ_i. From Equation (6), increases in G can be decomposed into increases in across-group inequality and increases in within-group inequality.

2.3 De-composing inequality effects

The across-group Gini is always increasing with group dispersion. Formally, this effect can be evaluated by taking the derivative of the across-group Gini by the dispersion measure of one or both groups. The derivative is always positive; an increase in dispersion will always increase the across-group Gini coefficient.[10] Because the log-normal distribution has positive mass across the entire income range, there is always *some* overlap; this is why the across-group term depends on σ even for small dispersions.

Milanovic (2002, p. 82-83) discusses the relationship between group means, group dispersions and income overlaps. He shows that for the overlap to be small, groups must either be very homogeneous internally (low within-group dispersion), or their mean incomes must be very far apart. Equation (6) allows for a formal

[10]The derivative with respect to $\sigma_i^2 + \sigma_j^2$ is

$$\frac{\partial G_A}{\partial\sqrt{\sigma_i^2+\sigma_j^2}} = \frac{y_j}{\bar{y}}\phi\left(\frac{\log\left(\frac{y_j}{y_i}\right)}{\sqrt{\sigma_i^2+\sigma_j^2}} + \frac{\sqrt{\sigma_i^2+\sigma_j^2}}{2}\right) + \frac{y_i}{\bar{y}}\phi\left(\frac{\log\left(\frac{y_j}{y_i}\right)}{\sqrt{\sigma_i^2+\sigma_j^2}} - \frac{\sqrt{\sigma_i^2+\sigma_j^2}}{2}\right)$$

The derivative with respect to σ_i or $c_i = s_i/y_i$ can then be found by the chain rule; this will not change the sign.

discussion of this. Consider an increase in the dispersion of group j, and the mean pairwise income difference between individuals in group j and (the poorer) group i. If the groups did not overlap; there would be no change; the lower distance resulting from a decrease in the income of the poorer individuals would be exactly offset by the increase in the income of the richer individuals, as the mean of group j is unchanged. With overlap, however, some of the poorest j-individuals are moving *away* from the richest i-individuals; the overlap makes the effect of increased dispersion greater. The degree of overlap is again influenced by the distance between groups ($\log\left(\frac{y_j}{y_i}\right)$) and the dispersion level ($\sigma_i^2 + \sigma_j^2$). This means that lower distance between groups increases the effect on the overlap term from increasing dispersion; groups that are close will have larger overlaps. The effect of changing dispersion is smaller for very large or very small dispersions; this reflects the bounding of the Gini coefficient to be between 0 and 1.

The last term in (6) is the sum of within-group Gini coefficients; a weighted sum of the Gini coefficients for log-normal distributions as reported by Aitchison & Brown (1957). It is straightforward to see that the within-group Gini increases with dispersion. As within-group pairs constitute a relatively small part of all possible pairs, the weights are low; for small groups, the squaring of the population share means that the resulting inequality contribution is low.

Returning to the aggregate Gini coefficient, it is useful to verify that Equation (6) takes on familiar values at the extremes of dispersion. First, consider a situation where within-group dispersion approaches zero: $\sigma_i \to 0$; in that case, the across-group Gini collapses to the between-group Gini (7) as both Φ functions are evaluated at plus infinity. Similarly, we can consider a situation where dispersion approaches infinity; in that case, as $\sigma \to \infty$, the Φ evaluations on y_j and y_i are evaluated at plus and minus infinity, respectively. The Gini coefficient approaches $\sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j y_i/\bar{y}$, which sums to 1; full inequality.

2.4 Finding within-group dispersions

From the discussion above we now know that when group distributions are log-normal, we can calculate aggregate and composite inequality measures in closed form, given group sizes, means and standard deviations. The standard deviations are not in the social tables. Because of this, we have to make a case for the "correct" level of within-group dispersion in each case to calculate aggregate inequality.

The following paragraphs discuss three possible ways of inferring reasonable ranges for inequality within groups. We will describe dispersion within each group in terms of coefficients of variation, $c_i = s_i/y_i$. In Section 3 below, a wide range of dispersion parameters will be examined.

2.4.1 Within-group dispersion in pre-industrial societies

From the three pre-industrial distributions discussed in Section 1.3, one can calculate the magnitude of dispersion directly. The means (across groups) of three inequality coefficients are reported in Table 4: the coefficient of variation c, the variance of log income (or wealth) $\tilde{\sigma}^2$, and the within-group Gini coefficient G_i.

Population	Mean c	Mean $\tilde{\sigma}^2$	Mean G_i
Tuscany, 1427 (Wealth)	2.12	2.03	0.64
Bihar, 1807	0.75	0.36	0.34
Norway, 1868	0.48	0.21	0.20

Table 4: Within-group inequality in pre-industrial societies

As explained above, all of these groups have some peculiarities in terms of the data. In the case of Tuscany, the data is on wealth, not distribution. In the case of Norway, the income data is only for the upper third of the distribution. And for Bihar, we only have two sectors. Moreover, some of the Bihar households are very large, which potentially leads to an underestimation of inequality as we have no within-household distribution data.

The limitations in the Bihar and Norway data can help explain why the measured inequality levels are so much lower than in Tuscany. On the other hand, the values for Tuscany are probably too high, as they concern wealth inequality, not income inequality. As all these three pre-industrial distributions have some limitation in terms of coverage, it will be useful to also look at other ways of inferring information about within-group dispersion.

2.4.2 Well-apportioned groups

In addition to inference from the three pre-industrial data sets, we can extrapolate inequality information from the distribution of income across groups to the distribution within groups. A possible approach is to say that groups should be "well-apportioned"; for a group to have a separate identity when tabulating incomes, the differences within the group should be less than the differences across the groups. This can be operationalized by requiring that the weighted sum of within-group Ginis not being larger than the between-group Gini.

The maximal level of dispersion consistent with this well-apportionment assumption will be denoted c_w; it will differ across societies, as it is derived from the group means and sizes. To calculate c_w, insert for the definition of σ (2) and the dispersion structure in the expression for within-group inequality in (6), and equate the average within-group dispersion with the between-group Gini.

The standard deviation of logs becomes $\sigma_w = \sqrt{2}\Phi^{-1}\left(\frac{G_B+1}{2}\right)$. Inserted in (5), we get the expression for the upper bound on the Gini coefficient consistent with well-apportioned groups:

$$G_{\text{“well-apportioned”}} = \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \frac{y_i}{\bar{y}} \left[2\Phi\left(\Phi^{-1}\left(\frac{G_B+1}{2}\right) + \frac{\log\left(\frac{y_i}{y_j}\right)}{2\Phi^{-1}\left(\frac{G_B+1}{2}\right)} \right) - 1 \right] \tag{8}$$

where G_B is given by Equation (7); that is, the expression depends only on the means and group sizes in the original data. For a simple back-of-the envelope calculation of inequality comparison across societies, Equation (8) is a good candidate. The dispersion c_w makes the within-group Gini for each group equal to the between-group Gini of the population. It can be seen as an upper bound of dispersion by making the following claim: if within-group dispersion was really bigger than c_w, the compiler of the table would not have chosen the groups in this way, as they do not add to the "structuring" of information about the society. In addition, this assumption allows for the coefficient of variation within groups to vary across societies.

2.4.3 Within-group dispersion in modern societies

Modern census or other survey data often include information on income, as well as several characteristics that makes it possible to group the population into "social classes" corresponding to the social tables. Using data from the International Integrated Public Use Microdata Series (Minnesota Population Center, 2010), the coefficient of variation of income can be calculated for groups based on occupation, industry and employment class. The result of such a procedure on nine countries is outlined in the online Appendix.[11]

The median within-group coefficient of variation is between 0.7 (Canada, 1981) and 4.8 (Mexico, 2000), with most being around 1. If we pool all group definitions and countries together, 25 per cent of c-coefficients are lower than 1 and 26 per cent are higher than 2. There is no clear relationship between development status and dispersion, though the groupings by "employment class" consistently yield higher dispersions than the other two groupings.

If the dispersion of income c within a group was correlated with the level of income, we would have to take account for this in our assumptions on dispersion. However, this does not appear to be the case. Running the regression $c_i = \alpha + \beta y_i$ for each modern sample separately, β is only significantly different from zero in a small minority of cases. Hence, it will be assumed that coefficients of variations are constant across groups; that standard deviations are proportional to group income. Similar regressions on the relationship between within-group dispersion and the number of groups on the country level finds no significant results, suggesting that the number of groups does not drive variations in within-group inequality.[12]

The combination of evidence from pre-industrial and modern societies, as well as the assumption of well-apportioned groups, guides the choice of coefficients of variation that will be used to re-evaluate the social tables.

[11] The countries for which the required data was available are Brazil, Canada, Colombia, Mexico, Panama, Puerto Rico, South Africa, United States and Venezuela. Observations are spaced between 1970 and 2007.

[12] Details on these regressions are provided in the online Appendix.

3 Re-evaluating pre-industrial inequality

With the methodology in place, pre-industrial inequality can be re-evaluated using the social table data compiled by Milanovic *et al.*. The overall level of inequality goes up by a large amount when within-group inequality is accounted for. In addition, changing dispersion also affects how we rank the various societies in terms of inequality.

Seven different sets of assumptions on within-group dispersion will be illustrated. The first and second set are the measures used by Milanovic *et al.*. Their "Gini1" assumes no within-group inequality — this is the "point distributions" discussed above — and is equal to the between-group Gini coefficient.[13] The "Gini2" variable is the inequality associated with within-group inequality and perfect group sorting, for given group interval borders, as described by Kakwani (1980, chap. 6). While Gini1 corresponds to $c = 0$, Gini2 does not map into the methodology used in this paper.

For the groupwise log-normal distributions, the coefficient of variation will be assumed constant across groups.[14] The values for c shown here will be $0.1, 0.5, 1$ and 2, covering most of the range discussed above. There will also be an assumption set with "well-apportioned" groups, where the within-group Gini coefficients are equal to the between-group coefficients. These differ between populations, as the estimates are calculated from group means and sizes, but are still constant across groups within each population.[15] The assumption sets used are summarized in Table 5.

3.1 The level of inequality in pre-industrial societies

The Gini coefficients increase significantly when within-group dispersion is accounted for. Figure 2 shows how the calculated Gini coefficients are sensitive to assumptions on within-group dispersion. The Gini estimates used by Milanovic *et al.* ("Gini1" and "Gini2") span only a small range of the possible values. Even

[13] The between-group Gini, G_B, can be calculated by Equation (7).

[14] Most results hold up to other linear relationships between s_i and y_i. This is detailed in the Online Appendix.

[15] See Equation (8) for the calculation of the well-apportioned groups.

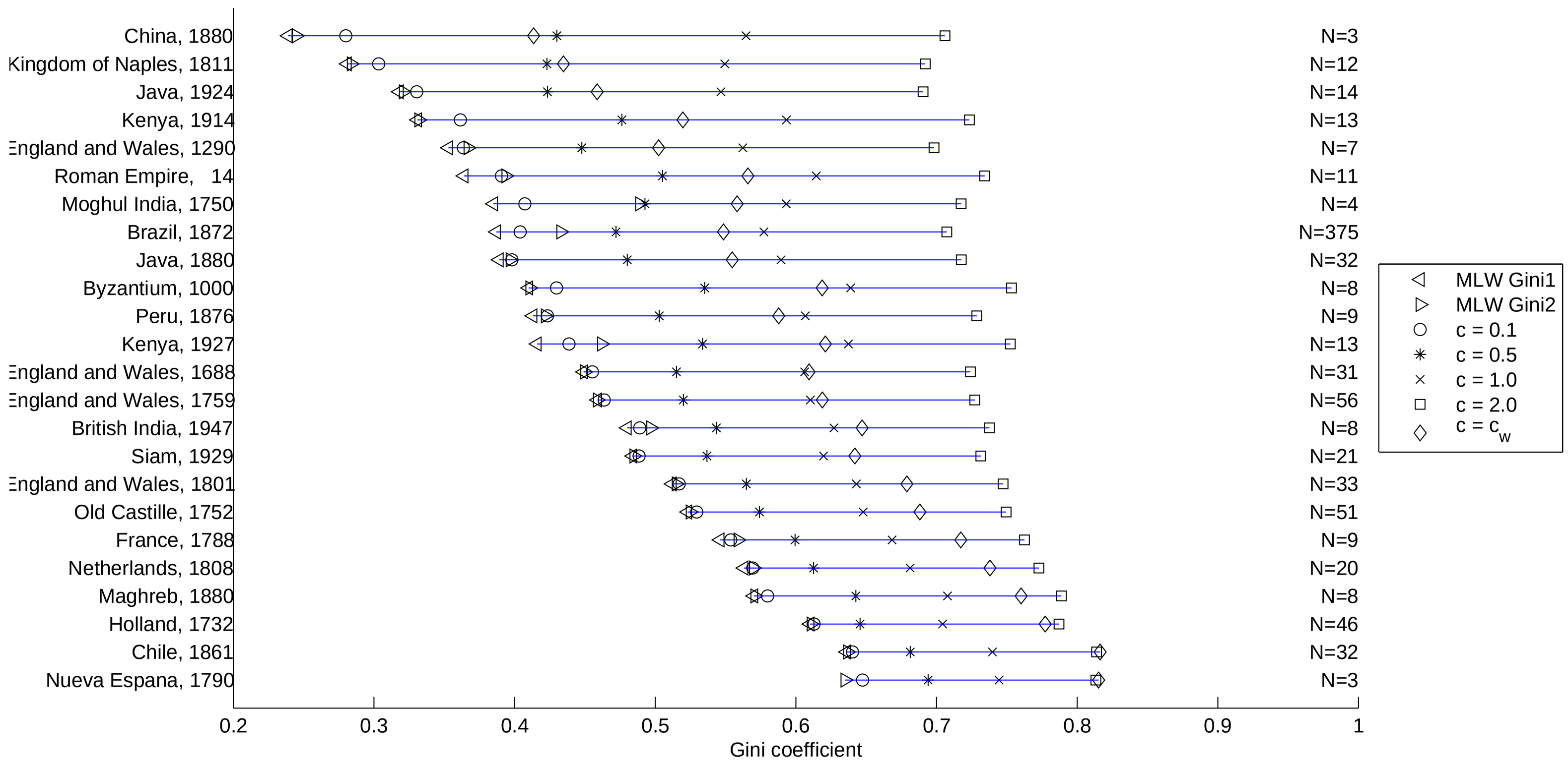


Figure 2: Comparison of Gini coefficients for the seven assumption sets

#	Within-group dispersion	Var. coeff c	Var. of log $\sigma^2 = \log(1+c^2)$	Gini within groups $G_i = 2\Phi(\sigma/\sqrt{2}) - 1$
1	None (MLW "Gini1")	0	0	0
2	Perfect sorting (MLW "Gini2")	-	-	-
3	Very low	0.1	0.01	0.06
4	Low	0.5	0.22	0.26
5	Intermediate	1	0.69	0.44
6	High	2	1.61	0.63
7	"Well-apportioned"	c_w	-	-

Table 5: Assumptions on within-group dispersions

the low coefficient of variance assumption of $c = 0.1$ gives higher Gini estimates for all but eight populations; increasing c to 0.35 leaves only Moghul India with higher Gini2. Like other populations with few groups, Moghul India has a large group containing the majority of the population; unlike the other populations, however, this group is not the poorest, and the income distance to the richer and poorer groups is relatively high. This allows for high inequality while preserving the assumption of no overlap. In the terms of Figure 1, the data points for Moghul India allow a large distance between the solid and dotted line, while for the other populations, this space is very small.

From Section 2.4.3 above, we know that the most coherent modern-day social groups have coefficients of income variations between .5 and 1. Using the still low value of $c = 0.5$, the calculated Gini coefficients for all the pre-industrial populations are higher than the Gini2 value. Further increasing within-group dispersion to $c = 2$, all Gini coefficients are higher than 0.7; very high inequality by any standard.

There is some change in sorting as c increases. At $c = 0.5$, around 7 per cent of all pairwise comparisons of societies change; at $c = 2$ this number has increased to 13 per cent. Above $c = 2$ the re-shuffling does not increase much more.[16] For the societies with higher between-group inequality, that is, the lower half of Figure 2, the sorting of societies is almost perfectly preserved — for example, by all measures, England and Wales in 1759 was just a little bit more unequal than in 1688. Hence, we can conclude that while the level of inequality is very sensitive

[16]For comparison, the expected change in pairwise sorting for random data sets is around 1/2 (50%).

to assumptions on within-group dispersions, the ranking is not.

With a large within-group dispersion measure, $c = 2$, calculated Gini coefficients are in some cases more than twice as large as the benchmark values. If inequality in these societies was this high, the value of the social tables data is low, as we would expect there to be variation in dispersion between populations, making it harder to rank the societies with respect to each other.

It could be a source of concern if the Gini coefficient of a population was highly dependent on the number of groups in that population. On the one hand, a high number of recorded groups could reflect a highly stratified society with corresponding inequality. On the other hand, we must assume that the number of recorded groups also reflects some pragmatism on the associated (often contemporary) researcher's part, with respect to how much data it is possible to collect. In any case, there is not a high correlation between the number of groups and the Gini estimates; for all estimation sets, linear OLS regression does not yield a significant slope parameter.[17]

To sum up, there are two main messages from Figure 2. First, the level of pre-industrial Gini coefficients is in general sensitive to assumptions on within-group dispersions. Second, the ordering of societies with respect to each other experiences some changes, but only around 10% of all compared pairs change order when the coefficient of variation within groups goes from 0 to 1.

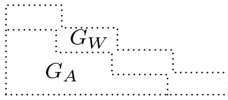
3.2 The contributions of subgroups to inequality

As discussed in the previous section, the increase in inequality comes both from inequality within and across groups. Using Equation (6), we can look at the contributions of group pairs to inequality. From each pair of groups, we get the weighted sum of pairwise income differences between individuals of the groups. As an example, consider again the social table for Byzantium, AD 1000, as given in Table 1. A Gini decomposition based on group pairs, with within-group dispersion at $c = 1$, is given in Table 6.

The upper panel shows the entire Gini coefficient. The diagonal is the within-

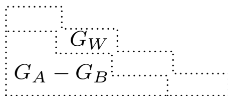
[17]This holds regardless of whether Brazil 1872, with 375 groups, is included in the regression. See the Online Appendix.

All Gini components ($G_A + G_W$)



	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
$j=1$	3.4							
$j=2$	0.4	0.0						
$j=3$	10.1	0.5	7.3					
$j=4$	0.8	0.0	1.2	0.0				
$j=5$	0.3	0.0	0.4	0.0	0.0			
$j=6$	3.1	0.2	4.4	0.2	0.1	0.2		
$j=7$	1.3	0.1	1.8	0.1	0.0	0.1	0.0	
$j=8$	10.3	0.6	14.5	0.8	0.3	0.9	0.3	0.1

"Within" and "overlap" terms ($G_A - G_B + G_W$)



	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
$j=1$	3.4							
$j=2$	0.4	0.0						
$j=3$	9.1	0.5	7.3					
$j=4$	0.4	0.0	0.6	0.0				
$j=5$	0.1	0.0	0.2	0.0	0.0			
$j=6$	0.1	0.0	0.2	0.0	0.0	0.2		
$j=7$	0.0	0.0	0.0	0.0	0.0	0.1	0.0	
$j=8$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

Table 6: Example of group pair contributions, Byzantium, AD 1000.

group Gini components; these would all be zero if there was no within-group dispersion. The other cells in the upper panel are the across-group components. Because groups are weighted by products of group sizes and incomes, small groups only add to inequality if differences *between* groups are very big. The lower row ($j = 8$) gives the contributions from the "nobility" group with very high income; because the difference from other groups is so big, interactions with this group contribute greatly to inequality. The most sizable contributions come from the interaction of the very small, very rich mobility group ($j = 8$) with the two poor, very big tenant and farmer groups ($i = 1, i = 3$). The sum of all the cells in the upper panel is the total Gini coefficient for this population, given a within-group coefficient of variation of 1.

Most of the large effects from group income differences come from the differences between group means, and are as such contained in the between-group Gini (G_B). The lower panel subtracts the between-group components,[18] giving the additions to inequality that arise solely from within-group dispersions.

When the between-group inequality is subtracted, nearly all contributions to inequality from the upper groups disappear. Within-group Gini coefficients, in particular for $i = 1$ and $i = 3$, the largest groups, contribute a total of 11 Gini points to the total Gini. [19] In this case, however, the across-group contribution is even more important. Inequality across farmers (group 1) and tenants (group 3) - large groups that have means close together - is particularly evident. This combination adds 9.1 points to a total Gini coefficient of 64 — nearly half the increase from the between-group Gini of 41. This highlights the restriction an assumption of perfect sorting places on inequality. As the means are so close, any perfectly sorted within-group distribution would have both these groups compressed over a very short income range.

Table 7 shows the decomposition of the increase in inequality for all the societies. For no within-group dispersion ($c = 0$), by construction, the within-group Gini is zero and the across-group component is equal to the between-group component. As c increases, both components go up; with many groups, more of the

[18] G_B is given in Equation (7).

[19] Throughout the text, Gini coefficients will be scaled to be between 0 and 100; a "Gini point" refers to a change of 1 in this measure.

		$c = 0$		$c = 0.5$		$c = c_w$	
		G_A	G_W	G_A	G_W	G_A	G_W
Roman Empire, 14	$(N = 11)$	36	0	38	12	40	17
Byzantium, 1000	$(N = 8)$	41	0	47	7	52	10
England and Wales, 1290	$(N = 7)$	35	0	40	5	44	6
England and Wales, 1688	$(N = 31)$	45	0	50	2	58	3
Holland, 1732	$(N = 46)$	61	0	64	1	76	2
Moghul India, 1750	$(N = 4)$	39	0	39	10	41	15
England and Wales, 1759	$(N = 56)$	46	0	51	1	60	2
Old Castille, 1752	$(N = 51)$	52	0	56	1	66	3
France, 1788	$(N = 9)$	55	0	57	3	66	6
Nueva Espana, 1790	$(N = 3)$	63	0	64	6	67	14
England and Wales, 1801	$(N = 33)$	51	0	55	2	64	4
Netherlands, 1808	$(N = 20)$	56	0	59	3	68	6
Kingdom of Naples, 1811	$(N = 12)$	28	0	40	2	41	2
Chile, 1861	$(N = 32)$	64	0	67	2	78	4
Brazil, 1872	$(N = 375)$	40	0	46	1	53	2
Peru, 1876	$(N = 9)$	41	0	46	4	52	7
China, 1880	$(N = 3)$	24	0	24	19	24	17
Java, 1880	$(N = 32)$	39	0	44	4	50	5
Maghreb, 1880	$(N = 8)$	57	0	60	4	67	9
Kenya, 1914	$(N = 13)$	33	0	34	14	34	18
Java, 1924	$(N = 14)$	32	0	39	3	42	4
Kenya, 1927	$(N = 13)$	42	0	43	10	46	16
Siam, 1929	$(N = 21)$	48	0	52	1	62	3
British India, 1947	$(N = 8)$	48	0	50	4	58	7

Table 7: Gini coefficients decomposed for different levels of within-group dispersion

increase is in across-group inequality, as more of the possible pairs of people are in separate groups. Some populations are clear outliers. For example, the social table for China has nearly all the population in the poorest group, and hence the "within" term of this group accounts for nearly the entire increase in G for high c. For Chile, the difference between group means is so big that increasing within-group dispersion has a less pronounced effect on both components. And for Naples, where group means are close, nearly all the increasing inequality is from increases in the across-group component.

The contribution to inequality from the affluent groups

For the richer income groups of historical inequality data (the upper social classes), we often have more detailed information on group structures. Hence, imposing the log-normal distribution, with positive mass across the entire income spectrum and a left-skewed distribution, might be harder to accept for these groups.

However, these upper groups are typically small, and it turns out that the contribution to aggregate inequality from dispersion within these groups is also small. As an example, consider the decomposition illustration of Table 6.

As is seen in the left column of the upper panel, the contributions to overall Gini from the richest group ($j = 8$) are substantial, even though it only consists of one per cent of the total population. However, all of this contribution comes from the difference in group means, which is present before the within-group dispersion is introduced. If we remove the between-group inequality, and move to the lower panel, it is clear that the contribution of the upper group is very low. As there is almost no overlap with the other groups, and the population of the richest group is low, the contribution of the richest group to the increased dispersion is almost zero.

Similar exercises can be conducted for the other social tables. Counting the "inequality contribution" from a group as all terms in (6) that include the group, we can check how much the richer groups contribute to overall inequality. Taking as the threshold any groups with a mean income of more than five times the population mean, and using the assumptions of $c = 1$, the result of this accounting exercise shows that there are no large contributions by the rich groups.[20]Even for the cases where these groups make up a considerable size of the population (they are largest in France and New Spain), the contribution from these groups only make up a small factor of the inequality that is added by within-group dispersion. It follows that removing the assumption of log-normal distributions within groups for the richer groups would not significantly alter the results in this paper.

3.3 Introducing a subsistence minimum assumption

Log-normal distributions have positive mass across the entire positive income range. Hence, by assuming such distributions within groups, we postulate that many people are very poor. However, some positive income level needs to be fulfilled in order to survive - the *subsistence income*. If we believed that everyone, at all times, lived at or above subsistence, we would have to revise our assumptions on within-group distributions. Inequality-limiting subsistence is one the key messages of Milanovic *et al.* (2011). As an example, the mean income of "Agricultural day laborers and servants" in France 1788 was 312 PPP dollars a year. With subsistence income at 300 dollars (as assumed in their paper), most people in that group (covering 36 per cent of the population) must have had incomes very close to the mean.

There is no need to assume that the subsistence border holds with absolute certainty; indeed, there is ample historical evidence to suggest that large groups have been living below subsistence level for long periods of time. A notable example is given in Clark (2008, chapter 6), where the Malthusian period is described as a situation with "social mobility and the survival of the richest". In pre-industrial England, according to Clark, poor families on average did not replace their population, while rich families did; consequently, there was continuous social mobility downward. However, it is not unlikely that subsistence income plays *some* role in truncating income distributions at the bottom, and it is useful to see how the results presented would change if the income of everyone was above subsistence minimum. In order to explore the effect on inequality on imposing subsistence minima, the setup of Section 2 is altered in three ways, starting with log-normal

[20]The table is given in the Online Appendix.

distributions built on a coefficient of variation of 1.

The first two adjustments keep the same log-normal distributions, but alter them at the tails. For the first adjustment, denoted "Cut" in Table 8, any population below the subsistence minimum is simply shifted up to the subsistence minimum. This reduces inequality at the lower end, but skews group means, as the same group-wise log-normal distributions are kept for the rest of the population. The second adjustment, labeled "Cut, preserve mean", addresses this by also shifting the richest part of the population in each group down to a "group upper bound", in such a way as to keep group means at the pre-adjustment levels.

The final adjustment ("Shift") is of a different type. Instead of defining the log-normal distribution on the entire positive income scale (starting at 0), it is defined over the scale starting at y_{min}. This means that there is no population mass below y_{min}. In practice, this amounts to subtracting y_{min} from all group means before calculating the log-normal distributions, and then right-shifting these distributions by y_{min}.

For each of these three adjustments, the aggregate Gini coefficients are recalculated. The calculation is done using numerical methods, calculating all pairwise differences in a discrete (but very fine-grained) population space.[21] Subsistence incomes are taken from Table 2 of Milanovic *et al.* (2011); however, in many cases (denoted by an asterisk in the table) the mean income of the poorest group is lower than this subsistence level. In those cases subsistence minimum is set to the mean income of the poorest group.

An adjustment by minimum incomes does shift the Gini estimates down for several populations, while others are virtually unchanged. Three populations stand out with large corrections: Byzantium and the two Kenya observations. All of these three have rather low population mean incomes, making the minimum income more quantitatively important; the population mean in Kenya 1914 is only 50% above minimum. Here, the same subsistence income is used for all populations; one could argue that the subsistence level is lower in tropical areas. If subsistence income in Kenya is actually lower, the downward revision of the Gini coefficient would be less.

A strong downward change in the Gini is expected across the line, as assump-

[21]For a full description of this procedure, see the online Appendix.

	$y_{min}/\bar{y}$	Benchmark G $(c=1)$	Cut	Cut, preserve mean	Shift	Benchmark G_B $(c=0)$
Roman Empire, 14	0.48	61	55	45	47	36
Byzantium, 1000	0.56	64	55	42	44	41
England and Wales, 1290	0.47*	56	50	44	44	35
England and Wales, 1688	0.21*	61	59	58	57	45
Holland, 1732	0.07*	70	70	70	70	61
Moghul India, 1750	0.30*	59	56	54	53	39
England and Wales, 1759	0.17	61	60	60	59	46
Old Castille, 1752	0.07*	65	65	65	64	52
France, 1788	0.26	67	63	61	62	55
Nueva Espana, 1790	0.24*	74	71	68	69	63
England and Wales, 1801	0.11*	64	64	64	63	51
Netherlands, 1808	0.17	68	67	66	65	56
Kingdom of Naples, 1811	0.45	55	49	43	43	28
Chile, 1861	0.16*	74	73	72	71	64
Brazil, 1872	0.23*	58	56	55	54	40
Peru, 1876	0.33*	61	57	54	53	41
China, 1880	0.56	56	48	37	39	24
Java, 1880	0.31*	59	56	54	53	39
Maghreb, 1880	0.32*	71	66	62	63	57
Kenya, 1914	0.66*	59	48	34	34	33
Java, 1924	0.33	55	52	49	48	32
Kenya, 1927	0.53	64	55	44	48	42
Siam, 1929	0.18*	62	61	60	59	48
British India, 1947	0.23*	63	60	59	59	48

Table 8: The Gini coefficients under different assumptions on minimum incomes, with $c = 1$

tions of no population mass below minimum income correspond directly to assumptions of very low within-group inequality at the bottom of the income distribution. The fact that substantial inequality (inequality above G_B) remains even after such an extreme revision shows that group overlap always needs to be accounted for when using group data, even if one adheres strongly to limiting subsistence incomes.

4 Concluding discussion

This paper has shown that when accounting for within-group inequality in social tables, reported inequality rises by a large amount. The increase comes from both within- and across-group inequality, and is particularly important in the case where groups are large and have means that are close to each other.

The log-normal distribution as used in this paper has the advantage of admitting a closed-form expression for the Gini coefficient and allows for overlap between the group-specific income distributions. For distributions where we have more knowledge about individual groups, other types of distributions might be more appropriate. Beyond the discussion of top and bottom incomes above, this is left for future work.

With further research, we can expect to see more tabulations of income and wealth data from pre-industrial societies. For statistics of a social table format, where within-group dispersion is not given, this paper presents a straightforward, transparent way of calculating inequality. The method can also be useful for modern data. While nation-wide distribution data now exist for most countries, within-group data is frequently missing for subnational entities or social classes. The approach presented in this paper can be used in these cases, to put structure on and properly evaluate any type of incomplete data on income or wealth distributions.

A Appendix: Calculations of expressions

A.1 Calculation of Equation (5)

This section shows the derivation of Equation (5), using the definition of the Gini coefficient as the area below the Lorenz curve. The calculation is an extension of Aitchison & Brown (1957)'s one-group case, and makes use of some convenient properties of the log-normal distribution.

Denote the log-normal population density functions as $f(x; \mu_i, \sigma_i^2)$ and the corresponding CDF as $F(x; \mu_i, \sigma_i^2) = \int_0^x f(u, \mu_i, \sigma_i^2)\mathrm{d}u$. Throughout this section, without loss of generality, group means will be rescaled to population means; that is, the population mean is always 1.

First, as stated by Aitchison & Brown, Theorem 2.6, page 12

$$\frac{1}{y_i}\int_0^x uf(u; \mu_i, \sigma_i^2)\mathrm{d}u = \int_0^x f(u; \mu_i + \sigma_i^2, \sigma_i^2)\mathrm{d}u \tag{9}$$

where y_i is the group mean.

Secondly, from Aitchison & Brown, Corollary 2.2b, page 11

$$\int_0^\infty F(ax; \mu_1, \sigma_1^2)\mathrm{d}F(x; \mu_2, \sigma_2^2) = F(a; \mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) \tag{10}$$

Now consider a piecewise log-normal distribution, with the probability density function

$$g(x) = \sum_{i=1}^{N} p_i f(x; \mu_i, \sigma_i^2) \tag{11}$$

The Lorenz curve plots cumulative population against cumulative income. Letting both axes run over income x, cumulative population is $G(x) = \int_0^x g(u)\mathrm{d}u$ while cumulative income is $V(x) = \int_0^x ug(u)\mathrm{d}u$.

By (9), cumulative income is

$$V(x) = \int_0^x u \sum_{i=0}^{N} p_i f(u; \mu_i, \sigma_i^2) \mathrm{d}u \tag{12}$$

$$= \sum_{i=1}^{N} p_i y_i \left(\frac{1}{m_i} \int_0^x u f(u; \mu_i, \sigma_i^2) \mathrm{d}u \right) \tag{13}$$

$$= \sum_{i=1}^{N} p_i y_i \left(\int_0^x f(u; \mu_i + \sigma_i^2, \sigma_i^2) \mathrm{d}u \right) \tag{14}$$

$$= \sum_{i=1}^{N} p_i y_i F(x; \mu_i + \sigma_i^2, \sigma_i^2) \tag{15}$$

Denote the total area below the Lorenz curve as H. It can be expressed as

$$H = \int_0^\infty V(x) \mathrm{d} \left[G(x) \right] \tag{16}$$

$$= \int_0^\infty \sum_{i=1}^{N} \left[p_i y_i \left(F(x; \mu_i + \sigma_i^2, \sigma_i^2) \right) \right] \mathrm{d} \left[\sum_{j=1}^{N} \left(p_j F(x; \mu_j, \sigma_j^2) \right) \right] \tag{17}$$

$$= \sum_{i=1}^{N} \left(p_i y_i \int_0^\infty F(x; \mu_i + \sigma_i^2, \sigma_i^2) \mathrm{d} \left[\sum_{j=1}^{N} \left(p_j F(x; \mu_j, \sigma_j^2) \right) \right] \right) \tag{18}$$

Reordering and using (10) to get

$$H = \sum_{i=1}^{N} \left(p_i y_i \sum_{j=1}^{N} p_j \left(\int_0^\infty F(x; \mu_i + \sigma_i^2, \sigma_i^2) \mathrm{d} \left[F(x; \mu_j, \sigma_j^2) \right] \right) \right) \tag{19}$$

$$= \sum_{i=1}^{N} \left(p_i y_i \sum_{j=1}^{N} p_j \left(F(1; (\mu_i - \mu_j) + \sigma_i^2, \sigma_i^2 + \sigma_j^2) \right) \right) \tag{20}$$

$$= \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \left(F(1; (\mu_i - \mu_j) + \sigma_i^2, \sigma_i^2 + \sigma_j^2) \right) \right) \tag{21}$$

Letting F_N denote a normal distribution and Φ its standardized variant, this can further be written as

$$H = \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \left(F_N(0; (\mu_i - \mu_j) + \sigma_i^2, \sigma_i^2 + \sigma_j^2) \right) \right) \tag{22}$$

$$= \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \left(\Phi \left(\frac{0 - (\mu_i - \mu_j + \sigma_i^2)}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \right) \right) \tag{23}$$

$$= \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \Phi \left(\frac{-(\mu_i - \mu_j + \sigma_i^2)}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \right) \tag{24}$$

$$= 1 - \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \Phi \left(\frac{\mu_i - \mu_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \right) \tag{25}$$

Finally, by the definition of the Gini coefficient,

$$G = 1 - 2H \tag{26}$$

$$= 2 \sum_{i=1}^{N} \left(y_i \sum_{j=1}^{N} p_i p_j \Phi \left(\frac{\mu_i - \mu_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \right) - 1 \tag{27}$$

A.2 Calculation of c_w

This section outlines the calculation of c_w. First, consider the more general case, where the relationship between standard deviations and group means are

$$\frac{s_i}{\bar{y}} = \alpha \left(\frac{y_i}{\bar{y}} \right)^{\beta} \tag{28}$$

α_w is defined as the α that makes the average of within-group Gini coefficients equal to the between-group Gini coefficient.

From Equations (2) and (28), we get

$$\sigma = \sqrt{\log\left(1 + \alpha^2 (y_i/\bar{y})^{2\beta-2}\right)} \tag{29}$$

α_w is then defined by the α that makes the average within-group Gini coefficient (right-hand side below) equal to the between-group Gini coefficient (left-hand side below; calculated from y and p).

$$G_B = \sum_{i=1}^{N} p_i 2\Phi \left[\sqrt{\frac{1}{2} \log\left(1 + \alpha_w^2 (y_i/\bar{y})^{2\beta-2}\right)} \right] - 1 \tag{30}$$

This is solved numerically when $\beta \neq 1$.

Note that when $\beta = 1$, $c_w = \alpha_w$. In this case:

$$G_B = 2\Phi \left(\sqrt{\frac{1}{2} \log\left[1 + \alpha_w^2\right]} \right) - 1 \tag{31}$$

$$\alpha_w = c_w = \sqrt{\exp\left(2 \left[\Phi^{-1} \left(\frac{G_B + 1}{2} \right) \right]^2 \right) - 1} \tag{32}$$

For $\beta = 1$, all within-Ginis will be equal to the between-Gini. For $\beta \neq 1$, the average of within-Ginis will be equal to the between-Gini. This means that alternate averages (weighting by $y_i p_i^2$ instead of p_i, for example) would produce different values for α_w if $\beta \neq 1$, but do not matter for $\beta = 1$.

A.3 Calculating decile shares

When a fuller knowledge of the aggregate distribution is desirable, one can calculate percentile shares. In the following, ten groups will be assumed (deciles), but any partition is possible.

Let d be the vector of population lower bounds for the groups ($d = \{0, .1, .2, .3, ..., .9\}$). Without loss of generality, rescale income so that the population mean is 1.

The lower income bounds a are then found numerically by solving

$$\sum_{i=1}^{N} \big(p_i F(a_j; \mu_i, \sigma_i^2)\big) - d_j = 0; \tag{33}$$

for each $j \in \{1, 2, 3, ..., 10\}$. (Trivially, $a_1 = 0$). As F is strictly increasing, (33) only has one solution for each j.

The upper bounds b are then the lower bounds of the group above, $b_j = a_{j+1}$; $b_{10} = \infty$.

The mean income of each decile is

$$\delta_j = \sum_{i=1}^{N} p_i \int_{a_j}^{b_j} u f(u; \mu_i, \sigma_i^2) du \tag{34}$$

$$= \sum_{i=1}^{N} p_i \left(\int_0^{b_j} u f(u; \mu_i, \sigma_i^2) du - \int_0^{a_j} u f(u; \mu_i, \sigma_i^2) du \right) \tag{35}$$

From Equation (9) this equals

$$\delta_j = \sum_{i=1}^{N} p_i y_i \left[\int_0^{b_j} f(u; \mu_i + \sigma_i^2, \sigma_i^2) du - \int_0^{a_j} f(u; \mu_i + \sigma_i^2, \sigma_i^2) du \right] \tag{36}$$

$$= \sum_{i=1}^{N} p_i y_i \left[F(b_j; \mu_i + \sigma_i^2, \sigma_i^2) - F(a_j; \mu_i + \sigma_i^2, \sigma_i^2) \right] \tag{37}$$

From this, for each decile group j, we know the bounds (a_j, b_j) and the mean income δ_j.

References

Aitchison, J., & Brown, J. A. C. 1957. *The Lognormal Distribution, with special reference to its uses in economics*. 1st edn. Cambridge University Press.

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

Clark, Gregory. 2008. *A Farewell to Alms: A Brief Economic History of the World*. Princeton University Press.

Crow, Edwin L., & Shimizu, Kunio. 1987. *Lognormal Distributions: Theory and applications*. 1 edn. CRC Press.

Ebert, Udo. 2010. The decomposition of inequality reconsidered: Weakly decomposable measures. *Mathematical Social Sciences*, **60**(2), 94–103.

Gastwirth, Joseph L. 1972. The Estimation of the Lorenz Curve and Gini Index. *Review of Economics and Statistics*, **54**(3), 306–316.

Hoffman, Philip T., Jacks, David S., Levin, Patricia A., & Lindert, Peter H. 2002. Real Inequality in Europe since 1500. *Journal of Economic History*, **62**(2), 322–355.

Kakwani, Nanak. 1980. *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. A World Bank Research Publication.

Kuznets, Simon. 1955. Economic Growth and Income Inequality. *American Economic Review*, **45**(1), 1–28.

Lambert, Peter J., & Aronson, J. Richard. 1993. Inequality Decomposition Analysis and the Gini Coefficient Revisited. *Economic Journal*, **103**(420), 1221–1227.

Lindert, Peter H. 2000. Three centuries of inequality in Britain and America. *Handbook of Income Distribution*, **1**, 167–216.

Milanovic, Branko. 2002. True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *Economic Journal*, **112**(476), 51–92.

Milanovic, Branko. 2006. An estimate of average income and inequality in Byzantium around year 1000. *Review of Income and Wealth*, **52**(3), 449–470.

Milanovic, Branko, Lindert, Peter H., & Williamson, Jeffrey G. 2007. Measuring Ancient Inequality. *National Bureau of Economic Research Working Paper Series*, **13550**(Oct.).

Milanovic, Branko, Lindert, Peter H., & Williamson, Jeffery G. 2011. Pre-Industrial Inequality. *Economic Journal*, **121**(551), 255–272.

Minnesota Population Center. 2010. *Integrated Public Use Microdata Series, International: Version 6.0 [Machine-readable database].*

Van Zanden, Jan L. 1995. Tracing the Beginning of the Kuznets Curve: Western Europe during the Early Modern Period. *Economic History Review*, **48**(4), 643–664.

Yitzhaki, Shlomo, & Lerman, Robert I. 1991. Income stratification and income inequality. *Review of Income and Wealth*, **37**(3), 313–329.

Young, Alwyn. 2011. The Gini Coefficient for a Mixture of Ln-Normal Populations. *mimeo, London School of Economics*, Dec.

Online appendix to **Inequality in the very long run: inferring inequality from data on social groups** by Jørgen Modalsli

Published in *Journal of Economic Inequality*, doi: 10.1007/s10888-014-9279-6

June 27, 2014

B Appendix: Data

B.1 Pre-industrial inequality data

These three sources are used in Section 1.3.[22]

Tuscany, 1427: The *Catasto* gives wealth information for the entire population of Florence and the immediate surroundings; 85.4% of households (87.8% of the population) are reported as having positive wealth. The population are divided into 97 different occupational groups, some with only one household. There are 9780 households and a total population of 38269. MLW use income data from other sources to assess inequality; this does not give a source of within-occupation variation and hence cannot be used here. For this reason, the within-group analysis is used only on wealth inequality. In keeping with MLW, the population with "no occupation given" are grouped together (they define it as "predominantly rural"). **Source:** Data downloaded from `http://www.stg.brown.edu/projects/catasto`. Full citation: *Online Catasto of 1427. Version 1.3. Edited by David Herlihy, Christiane Klapisch-Zuber, R. Burr Litchfield and Anthony Molho. [Machine readable data file based on D. Herlihy and C. Klapisch-Zuber, Census and Property Survey of Florentine Domains in the Province of Tuscany, 1427-1480.] Florentine Renaissance Resources/STG: Brown University, Providence, R.I., 2002.*

Bihar, 1807: Expenditure is given as household total; expenditure per capita is found by dividing by household size. Both expenditure and household size are given as (narrow) intervals; a mean value has been used for both in the calculation.

[22]The first two are also used (aggregately) by Milanovic *et al.* (2011); when referring to that data set, their tabulations and calculations are used rather than the original sources. These are documented at `http://gpih.ucdavis.edu/Distribution.htm`.

The expenditure data is reported separately for 17 districts, one of which is Patna city. Here, all other districts are grouped together as "rural". **Source:** Montgomery Martin: "The history, antiquities, topography, and statistics of Eastern India, Volume I: Behar (Patna City) and Shahabad". London, W. H. Allen and Co., 1838. Appendix pages 6 and 7.

Norway, 1868: The source is a report published by the Department of Justice listing incomes for all males above 25 years of age with income above 100SPD (400 kr) and not belonging to the servant class. The total number of persons tabulated above the lower threshold amounts to 37.7% of the above-25 male population in the census of 1865; hence, around the upper third of the income distribution is covered (allowing for some population growth 1865-1868). This means that for the more prestigous occupations, coverage is near-complete, while for other occupations, the report only gives the upper tail of the distribution. There are a total of 26 occupations listed, divided into five income categories (100, 100-150, 150-200, 200-250, 250+). When calculating Gini coefficients, group mean incomes of 100, 125, 175, 225 and 400 are assumed; when fitting CDF functions, the four group borders are used. **Source:** "Tabeller til oplysning om stemmerets- indtægts- og skatteforholdene i Norge i aaret 1868". Udgivet av Justits-Departementet. Christiania, 1871. Table II, page 3.

Lognormality of pre-industrial distributions

As mentioned in the main text, a proper test of lognormality is hard to conduct on the pre-industrial data, as it is still "bracketed". For Bihar, the population is still grouped within households, some of the households are very large and it is implicitly assumed (both here and in MLW) that there is a perfectly egalitarian distribution within households. For Tuscany, the households are smaller but the problem exists to some extent. In addition, some occupation groups are very small. For Norway, there is only five bracketed income groups.

However, for the bracketed Norwegian data we can check how well the bracket borders (which correspond to well-defined points on the cumulative distribution function) correspond to an idealized lognormal distribution. This is done by a variant of a QQ-plot, plotting the idealized cumulative densities against the empirical densities for each group. The same exercise is done for the other two populations, but with the caveat of the undecided within-household distributions.

The plots are shown in Figure B.1.[23] For Tuscany, the degree of fit varies between groups (only groups with more than 200 households are shown). For Bihar, the urban distribution has a good fit while the rural has not; this might be related to the fact that there are more large households (and hence a less smooth dataset) in the rural area.

For Norway, the fit is very good for the more prestigous occupations, such as the high-level civil servants. For the other occupations, such as the workers, the problem of missing coverage of the lower part of the income distribution makes the fit somewhat worse.

[23]The Tuscan groups correspond are the ones numbered 0, 13, 21, 24, 29, 46 and 61 (in that order) in the original source, where a fuller description (with Italian-language titles) is given. The Norwegian groups are also presented in the same order as in the Norwegian-language source.

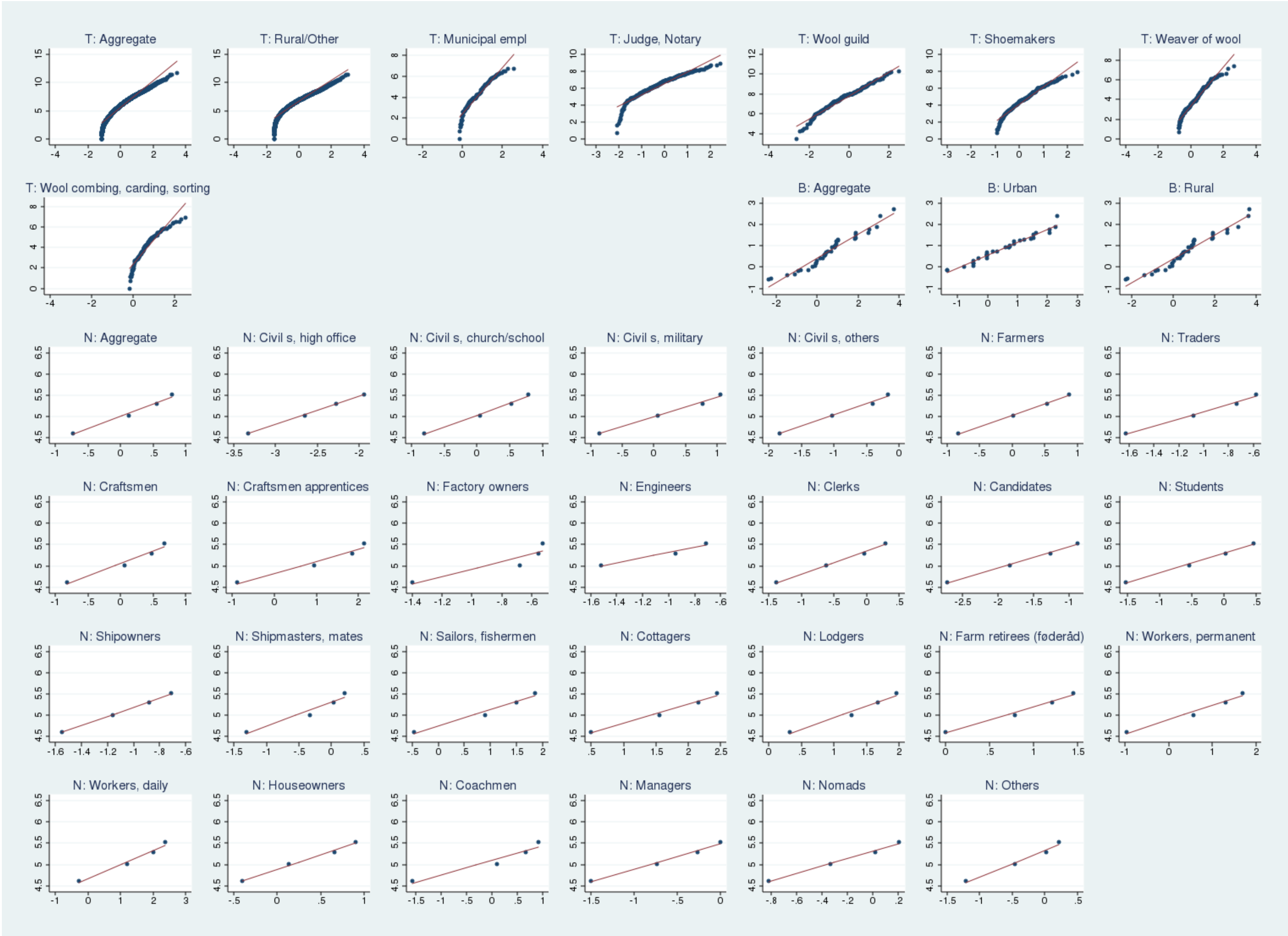


Figure B.1: Fit of lognormal CDF to aggregate and groupwise data. Horizontal axis: Inverse normal CDF of cumulative population. Vertical axis: Log income of cumulative population. Legend: T=Tuscany 1427, B=Bihar 1807, N=Norway 1868

B.2 Modern inequality data

Data for nine developed and developing countries between 1970 and 2007 is used: Brazil, Canada, Colombia, Mexico, Panama, Puerto Rico, South Africa, United States and Venezuela. All country-years are listed in Table B.2. A summary of the group data is given in Table B.1.

Classification	Mean of $c_{\min}$	Mean of c_{median}	Mean of c_{max}	Mean # of groups
Occupation	1.0	1.3	3.1	9.4
Industry	0.9	1.5	2.9	13.9
Empl.classification	1.5	2.0	6.0	2.7
Empl.class (detailed)	1.1	1.7	6.1	5.8

Table B.1: Within-group inequality (coefficient of variation) in modern societies

The range of variation coefficients is not large. Comparing the dispersion in the most and least diverse groups, for less than half of the country-years is the former more than three times the latter. Moreover, the the mean and minimum of the dispersion of groups are quite similar.

B.2.1 The level of within-group dispersion

The underlying information for Section 2.4.3 is in Table B.2.

The data has been compiled by IPUMS (Minnesota Population Center, 2010). As reported by IPUMS, the statistical data was originally produced by

- Brazil: Institute of Geography and Statistics
- Canada: Statistics Canada
- Colombia: National Administrative Department of Statistics
- Mexico: National Institute of Statistics, Geography, and Informatics
- Panama: Census and Statistics Directorate
- Puerto Rico: U.S. Bureau of the Census
- South Africa: Statistics South Africa

	Occupation (ISCO)				Industry (IPUMS)				Empl.class. (IPUMS)				Empl.class. (IPUMS, detailed)			
Country and Year	$c_{\min}$	c_{median}	$c_{\max}$	N	$c_{\min}$	c_{median}	$c_{\max}$	N	$c_{\min}$	c_{median}	$c_{\max}$	N	$c_{\min}$	c_{median}	$c_{\max}$	N
Brazil 1970	0.8	1.1	1.8	10	1.0	1.4	1.9	15	1.5	2.0	23.9	3	1.1	1.4	23.9	6
Brazil 1980	1.0	1.3	6.3	10	0.8	1.6	6.2	15	1.9	3.4	34.6	3	0.7	2.3	34.6	8
Brazil 1991	1.1	1.5	3.8	10	1.2	2.0	3.8	15	1.9	2.5	20.5	3	1.0	2.1	20.5	11
Brazil 2000	1.2	1.7	11.1	10	0.8	2.1	10.6	15	1.7	4.8	7.4	3	0.8	3.5	7.2	7
Canada 1971	0.6	0.8	1.8	8	0.7	0.8	1.7	9	0.9	1.3	4.5	3	0.9	1.2	4.5	4
Canada 1981	0.7	0.7	1.4	8	0.6	0.8	1.4	14	0.8	1.1	2.6	3	0.8	1.0	2.6	4
Canada 1991	0.7	0.7	1.2	8	0.7	0.9	1.2	13	0.8	1.0	1.1	2	0.8	1.1	1.1	3
Canada 2001	0.6	0.8	1.0	9	0.6	0.8	1.1	13	0.8	0.9	1.1	2	0.8	0.9	1.1	3
Colombia 1973	1.1	1.7	3.0	9	1.3	2.0	3.0	11	2.0	2.1	2.1	2	1.6	1.8	4.1	5
Mexico 1995	0.8	1.2	2.6	9	0.7	1.8	2.6	12	1.5	2.7	9.0	3	1.0	2.0	9.0	5
Mexico 2000	1.7	3.3	9.0	10	1.5	3.3	10.2	15	3.7	4.5	13.5	3	2.6	4.8	13.5	5
Panama 1980	0.8	1.1	8.7	9	0.7	1.2	4.3	14	1.5	2.8	6.2	3	0.9	2.0	7.4	5
Panama 1990	0.8	1.2	3.7	10	0.7	1.4	2.9	14	1.3	2.7	5.0	3	0.8	1.7	5.0	5
Puerto Rico 1970	0.9	1.0	1.4	9	0.9	1.2	1.4	11	1.2	1.3	1.4	2	0.9	1.3	1.4	3
Puerto Rico 1980	0.9	1.0	1.4	9	0.8	1.1	1.4	14	1.1	1.2	1.3	2	0.9	1.1	1.4	6
Puerto Rico 1990	1.0	1.1	1.3	9	0.7	1.2	1.4	14	1.1	1.3	1.4	2	0.9	1.2	1.4	7
Puerto Rico 2000	1.6	2.0	3.7	9	0.9	2.1	4.1	14	1.9	2.3	4.5	3	1.4	2.0	4.5	8
Puerto Rico 2005	1.0	1.1	1.5	8	0.9	1.2	1.6	12	1.2	1.4	1.6	2	0.8	1.1	1.7	7
South Africa 1996	0.9	1.1	2.4	9	0.8	1.4	2.5	15	1.4	1.5	1.5	3	1.3	1.4	1.7	4
South Africa 2001	1.8	2.4	3.7	9	1.8	2.6	4.0	15	2.4	2.6	2.8	2	2.4	2.7	3.3	4
South Africa 2007	1.7	2.3	4.2	10	1.5	2.5	3.7	14	2.3	2.5	2.5	3	2.3	2.5	2.5	3
United States 1960	0.6	0.9	1.3	10	0.6	1.1	1.5	15	1.0	1.2	4.3	3	0.8	1.1	4.3	4
United States 1970	0.6	0.9	1.3	10	0.7	1.1	1.5	15	1.0	1.1	3.2	3	0.8	0.9	3.2	7
United States 1980	0.7	0.9	1.2	10	0.6	1.0	1.3	15	0.9	1.1	2.5	3	0.8	0.9	2.5	7
United States 1990	0.7	1.0	1.3	10	0.6	1.2	1.4	15	1.1	1.3	2.4	3	0.7	1.1	2.4	8
United States 2000	0.9	1.2	1.5	10	0.8	1.3	1.8	15	1.3	1.5	2.5	3	0.8	1.2	2.5	8
United States 2005	0.7	1.0	1.4	10	0.7	1.2	1.7	15	1.1	1.4	2.0	3	0.8	1.1	2.0	8
Venezuela 2001	0.9	1.1	2.6	10	0.9	1.2	1.8	15	1.2	1.4	1.6	3	0.9	1.2	1.4	6

Table B.2: Modern within-group inequality. Source data: IPUMS

- United States: Bureau of the Census
- Venezuela: National Institute of Statistics

B.2.2 The distribution of within-group dispersion: c and the number of groups

To rule out a systematic relationship between within-group dispersion and the number of groups, we can regress the average dispersion within a sample (that is, a country-year-classification set) on the number of groups in the same sample. Denoting as $\bar{c}_j$ the average coefficient of variation over, for example, occupation groups in Brazil in 1970, and N_j as the number of such occupation groups, we have the regression equation $\bar{c} = \alpha + \beta N$ for all the country-year combinations for a given classification. For all four classifications separately, as well as a pooled regression with and without classification dummies, the null hypothesis of $\beta = 0$ cannot be rejected at a 95% level.

The relationship between the number of groups and the average coefficient of variation is shown in Table B.3. For each sample (Country and Year), the coefficient of variation is calculated for each group and the average over these groups are then taken. This average is then regressed on the number of groups in each sample. As the table shows, for no classification is there a significant correlation at the 95% level.

B.2.3 The distribution of within-group dispersion: c and mean income

Tables B.4 and B.5 show the relationship between the coefficient of variance and the mean income of each group. Here the regression is done within each sample. We see that in most cases the 95%-interval covers zero, meaning the correlations are not significant; however, for the "industry" classification we have a substantial number of coefficients significantly different from zero (always on the negative side). For this reason, a robustness check is done where β is set to -0.3 instead of 0 as in the main text (using the log-log specification). The results from this robustness check are shown in Section C.1.

Dependent variable: Average coefficient of variation					
	(1)	(2)	(3)	(4)	(5)
Classification	Occ	Ind	Class	Class det.	All
Const	-8.212	-3.435	-4.993	2.031	2.683**
	(9.370)	(6.619)	(5.380)	(2.611)	(1.140)
Number of groups	1.108	0.408	3.230	0.171	0.412
	(0.996)	(0.472)	(1.949)	(0.430)	(0.302)
R^2	0.044	0.027	0.092	0.006	0.040
N	29	29	29	29	116

Table B.3: Lack of correlation between average coefficient of variation and number of groups. Regression (5) pools 1-4 and has dummies for each of the four classifications. {***,**,*}=significant at {99%,95%,90%} level (two-sided tests)

	Occupation		Industry		Empl. class		Emp.c.detailed	
Country and Year	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Brazil 1970	-0.05	(-0.23,0.13)	-0.10	(-0.24,0.04)	-16.36	(-45.97,13.26)	-2.04	(-7.65,3.58)
Brazil 1980	-0.37	(-1.51,0.77)	-0.68	(-1.78,0.43)	-24.04	(-110.00,58.50)	-1.33	(-7.28,4.62)
Brazil 1991	-0.24	(-0.84,0.36)	-0.71	(-1.10,-0.33)	-10.72	(-48.30,26.86)	-1.21	(-3.61,1.19)
Brazil 2000	-0.33	(-1.63,0.98)	-1.26	(-2.89,0.38)	-2.04	(-34.46,30.38)	-0.16	(-1.20,0.88)
Canada 1971	-0.27	(-0.86,0.31)	-1.11	(-1.48,-0.74)	-2.42	(-15.12,10.28)	-1.59	(-5.01,1.83)
Canada 1981	-0.36	(-0.91,0.19)	-0.68	(-0.94,-0.41)	-1.35	(-9.21,6.51)	-1.01	(-2.97,0.94)
Canada 1991	-0.35	(-0.84,0.14)	-0.56	(-0.80,-0.32)	0.92	(N=2)	0.30	(-3.81,4.41)
Canada 2001	-0.20	(-0.49,0.10)	-0.43	(-0.64,-0.23)	1.81	(N=2)	-0.08	(-4.12,3.97)
Colombia 1973	-0.29	(-0.69,0.11)	-0.43	(-0.97,0.12)	0.07	(N=2)	-0.63	(-2.24,0.97)
Mexico 1970	-1.33	(-2.67,0.01)	-4.85	(-6.15,-3.55)	-9.42	(-38.10,19.26)	-9.47	(-12.05,-6.88)
Mexico 1995	-0.25	(-0.84,0.33)	-0.66	(-1.16,-0.16)	-5.70	(-16.95,5.54)	-0.81	(-4.16,2.54)
Mexico 2000	-0.90	(-1.81,0.01)	-2.12	(-3.48,-0.75)	-7.45	(-16.98,2.07)	-1.97	(-5.66,1.72)
Panama 1980	-0.96	(-3.89,1.96)	0.27	(-0.97,1.52)	-1.42	(-57.29,54.45)	-2.09	(-7.79,3.60)
Panama 1990	-0.52	(-1.27,0.24)	-0.40	(-0.82,0.03)	-3.04	(-9.98,3.91)	-0.76	(-1.91,0.39)
Puerto Rico 1970	-0.11	(-0.29,0.06)	-0.11	(-0.40,0.18)	0.61	(N=2)	-0.24	(-15.54,15.07)
Puerto Rico 1980	-0.14	(-0.27,0.00)	-0.22	(-0.45,0.02)	0.77	(N=2)	0.06	(-0.45,0.57)
Puerto Rico 1990	-0.09	(-0.17,-0.01)	-0.26	(-0.48,-0.03)	0.64	(N=2)	0.17	(-0.42,0.76)
Puerto Rico 2000	-0.69	(-1.47,0.09)	-1.42	(-2.20,-0.63)	-2.10	(-20.29,16.10)	-1.44	(-3.21,0.32)
Puerto Rico 2005	-0.03	(-0.25,0.20)	-0.09	(-0.42,0.25)	1.34	(N=2)	-0.23	(-1.02,0.57)
South Africa 1996	-0.20	(-0.49,0.08)	-0.37	(-0.57,-0.16)	0.11	(-1.08,1.29)	-0.07	(-0.50,0.36)
South Africa 2001	-0.27	(-0.51,-0.03)	-0.45	(-0.73,-0.17)	-0.21	(N=2)	-0.34	(-0.82,0.14)
South Africa 2007	-0.41	(-0.86,0.03)	-0.49	(-0.78,-0.20)	-0.03	(-0.11,0.06)	0.00	(-0.16,0.16)
United States 1960	-0.25	(-0.54,0.03)	-0.65	(-0.94,-0.35)	-2.40	(-14.15,9.34)	-2.53	(-5.71,0.66)
United States 1970	-0.18	(-0.43,0.06)	-0.51	(-0.83,-0.18)	-1.41	(-12.21,9.40)	-0.69	(-1.72,0.33)
United States 1980	-0.24	(-0.53,0.06)	-0.56	(-0.76,-0.35)	-1.11	(-9.73,7.50)	-0.57	(-1.41,0.27)
United States 1990	-0.11	(-0.46,0.24)	-0.41	(-0.73,-0.09)	-0.87	(-8.66,6.92)	-0.42	(-1.20,0.35)
United States 2000	-0.11	(-0.54,0.32)	-0.61	(-0.99,-0.24)	-0.95	(-10.83,8.93)	-0.44	(-1.38,0.51)
United States 2005	-0.14	(-0.53,0.26)	-0.45	(-0.72,-0.19)	-0.76	(-9.21,7.69)	-0.46	(-1.26,0.34)
Venezuela 2001	0.62	(-0.05,1.30)	-0.18	(-0.50,0.13)	3.15	(N=2)	0.23	(-0.28,0.74)
All (with country-year dummies)	-0.14	(-0.57,0.29)	-0.08	(-0.71,0.56)	-3.53	(-5.73,-1.32)	-1.02	(-1.83,-0.21)

Table B.4: Coefficents and confidence intervals from the regression $c = \alpha + \beta y/\bar{y}$

	Occupation		Industry		Empl. class		Emp.c.detailed	
Country and Year	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Brazil 1970	-0.11	(-0.35,0.13)	-0.09	(-0.26,0.07)	-0.62	(-0.96,-0.27)	-0.56	(-0.88,-0.23)
Brazil 1980	-0.25	(-0.87,0.38)	-0.07	(-0.55,0.41)	-0.61	(-2.60,1.37)	-0.42	(-0.95,0.12)
Brazil 1991	-0.23	(-0.61,0.14)	-0.40	(-0.59,-0.21)	-0.50	(-1.53,0.53)	-0.41	(-0.67,-0.15)
Brazil 2000	-0.29	(-0.91,0.32)	-0.24	(-0.86,0.38)	-0.27	(-3.94,3.40)	-0.19	(-0.62,0.25)
Canada 1971	-0.41	(-1.04,0.21)	-1.02	(-1.41,-0.63)	-0.41	(-2.11,1.30)	-0.38	(-0.79,0.03)
Canada 1981	-0.49	(-1.12,0.14)	-0.66	(-0.96,-0.37)	-0.50	(-2.77,1.77)	-0.47	(-0.98,0.03)
Canada 1991	-0.46	(-0.99,0.08)	-0.54	(-0.80,-0.27)	1.11	(N=2)	0.45	(-4.90,5.79)
Canada 2001	-0.28	(-0.63,0.07)	-0.44	(-0.68,-0.21)	2.08	(N=2)	-0.10	(-5.24,5.03)
Colombia 1973	-0.24	(-0.54,0.07)	-0.26	(-0.57,0.05)	0.04	(N=2)	-0.33	(-0.78,0.12)
Mexico 1970	-0.32	(-0.64,0.00)	-0.42	(-0.46,-0.38)	-0.49	(-2.11,1.13)	-0.45	(-0.63,-0.27)
Mexico 1995	-0.29	(-0.85,0.26)	-0.57	(-1.00,-0.13)	-0.39	(-1.96,1.18)	-0.35	(-0.85,0.16)
Mexico 2000	-0.65	(-0.86,-0.44)	-0.73	(-1.12,-0.34)	-0.40	(-0.79,-0.01)	-0.40	(-0.63,-0.17)
Panama 1980	-0.43	(-1.47,0.61)	0.07	(-0.52,0.66)	-0.12	(-7.28,7.05)	-0.35	(-1.44,0.74)
Panama 1990	-0.41	(-0.76,-0.06)	-0.23	(-0.56,0.09)	-0.37	(-2.11,1.36)	-0.38	(-0.64,-0.13)
Puerto Rico 1970	-0.15	(-0.35,0.04)	-0.14	(-0.39,0.12)	0.63	(N=2)	-0.33	(-17.18,16.51)
Puerto Rico 1980	-0.21	(-0.35,-0.07)	-0.16	(-0.39,0.07)	0.84	(N=2)	0.10	(-0.53,0.73)
Puerto Rico 1990	-0.12	(-0.23,-0.02)	-0.26	(-0.52,0.01)	0.72	(N=2)	0.16	(-0.54,0.86)
Puerto Rico 2000	-0.45	(-0.80,-0.09)	-0.76	(-1.16,-0.37)	-0.63	(-5.09,3.82)	-0.64	(-1.23,-0.04)
Puerto Rico 2005	-0.02	(-0.25,0.22)	-0.07	(-0.38,0.23)	1.22	(N=2)	-0.27	(-1.21,0.68)
South Africa 1996	-0.35	(-0.69,-0.02)	-0.38	(-0.56,-0.19)	0.15	(-1.47,1.78)	-0.11	(-0.82,0.60)
South Africa 2001	-0.26	(-0.43,-0.10)	-0.21	(-0.37,-0.05)	-0.19	(N=2)	-0.28	(-0.55,0.00)
South Africa 2007	-0.28	(-0.54,-0.01)	-0.28	(-0.50,-0.07)	-0.01	(-0.05,0.04)	0.00	(-0.04,0.04)
United States 1960	-0.33	(-0.66,0.00)	-0.40	(-0.66,-0.15)	-0.49	(-1.98,1.00)	-0.53	(-1.04,-0.02)
United States 1970	-0.25	(-0.55,0.05)	-0.29	(-0.55,-0.03)	-0.55	(-3.07,1.97)	-0.54	(-0.88,-0.19)
United States 1980	-0.29	(-0.63,0.04)	-0.43	(-0.66,-0.20)	-0.56	(-3.43,2.31)	-0.54	(-0.93,-0.15)
United States 1990	-0.15	(-0.55,0.25)	-0.32	(-0.63,-0.02)	-0.42	(-3.25,2.42)	-0.43	(-0.91,0.05)
United States 2000	-0.12	(-0.52,0.28)	-0.43	(-0.71,-0.15)	-0.48	(-4.48,3.53)	-0.44	(-1.10,0.22)
United States 2005	-0.16	(-0.56,0.24)	-0.34	(-0.55,-0.12)	-0.44	(-4.53,3.65)	-0.48	(-1.11,0.15)
Venezuela 2001	0.27	(-0.32,0.87)	-0.12	(-0.40,0.16)	2.32	(N=2)	0.23	(-0.18,0.64)
All (with country-year dummies)	-0.16	(-0.30,-0.01)	-0.06	(-0.19,0.08)	-0.44	(-0.56,-0.31)	-0.39	(-0.49,-0.28)

Table B.5: Coefficents and confidence intervals from the regression $\log(c) = \alpha + \beta\log(y/\bar{y})$

C Appendix: Robustness checks

C.1 More general specification of variance structure

As noted in Footnote 14, a more general specification of the variance structure is $c_i = \alpha(y_i/\bar{y})^{\beta}$. The specification used in the main text — with the coefficient of variation constant — corresponds to $\beta = 0$. However, cases could be made for other relationships between group mean and group dispersion; that is, other values for β. Figure C.1 shows the results for $\beta = -0.3$. The results do not greatly differ from those in the main paper.

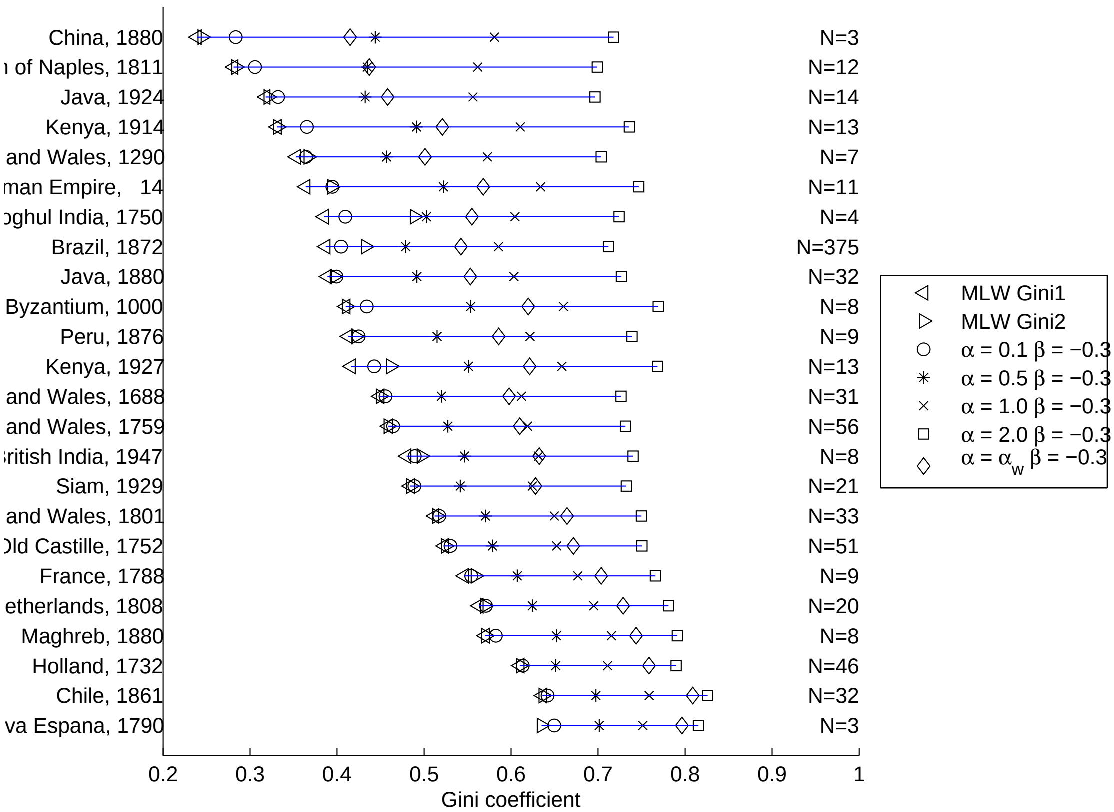


Figure C.1: Comparison of Gini coefficients for the seven assumption sets; $\beta = -0.3$

C.2 The inequality impact of upper groups

Results are shown in Table C.1.

As we count all terms except the within-group cells (the diagonal) as belonging to two groups, the sum of all these terms is not the overall Gini. In the table, the column "contributions of rich groups" includes all terms where the rich groups are at least one of the groups. For example, for the Byzantine example (Table 6), if the two richest groups were included as "rich", all cells that are part of the two rightmost columns and/or the two lowest rows would be included; the $(i = 7, j = 8)$ cell would not be counted twice toward the inequality contribution.

	N	N^r	n^r	G	G_W	G_W^r	G_R	G_R^r
Roman Empire, 14	11	6	0.04	61	21	0	4	0
Byzantium, 1000	8	1	0.01	64	11	0	12	0
England and Wales, 1290	7	1	0.04	56	8	0	13	0
England and Wales, 1688	31	8	0.02	61	3	0	13	0
Holland, 1732	46	15	0.04	70	1	0	8	1
Moghul India, 1750	4	1	0.01	59	17	0	3	0
England and Wales, 1759	56	13	0.02	61	2	0	14	0
Old Castille, 1752	51	9	0.04	65	2	0	10	1
France, 1788	9	2	0.10	67	5	1	7	1
Nueva Espana, 1790	3	1	0.10	74	10	3	1	0
England and Wales, 1801	33	8	0.04	64	3	0	10	1
Netherlands, 1808	20	10	0.03	68	4	0	7	0
Kingdom of Naples, 1811	12	1	0.01	55	3	0	24	0
Chile, 1861	32	6	0.05	74	3	0	8	1
Brazil, 1872	375	114	0.01	58	2	0	16	0
Peru, 1876	9	2	0.02	61	7	0	12	0
China, 1880	3	2	0.02	56	32	0	0	0
Java, 1880	32	22	0.01	59	6	0	14	0
Maghreb, 1880	8	1	0.01	71	7	0	7	0
Kenya, 1914	13	8	0.01	59	24	0	3	0
Java, 1924	14	2	0.01	55	6	0	17	0
Kenya, 1927	13	8	0.01	64	17	0	5	0
Siam, 1929	21	1	0.01	62	2	0	11	0
British India, 1947	8	2	0.01	63	7	0	8	0

Table C.1: Inequality contribution from the richest groups. Superscript r denotes contributions from groups with mean incomes more than five times greater than population mean

C.3 Adding subsistence income

This explains the numerical procedure used to calculate the values in Table 8, discussed in Section 3.3.

A population grid X of 50 000 points is constructed, with points spaced equally apart in logs (more points at the bottom). This combines the need for high accuracy at the bottom (where there is high "population density") with the need for covering large income ranges at the top (where density is lower, and one does not need as fine a grid). The grid runs from zero to 10 000 times the mean income of the richest group. A weight is assigned to each grid point corresponding to the inverse of the spacing of points.

Adjustments 1 and 2

The log-normal PDF is then calculated for each of these points for each group, and the distributions normalized to group sizes.

As y is already normalized so that the population mean is 1, subsistence income is found by inverting the number "mean income in terms of s" found in Table 2 of Milanovic *et al.* (2011). When the lowest income group has lower mean income than this subsistence group, the lowest group mean income will be chosen, subtracting 0.0001 (the scaling is population mean) to allow for some very small dispersion at the bottom group; this does not alter the results, but simplifies the calculation.

Then, for each population group, the total mass of everyone below subsistence income P is calculated, replacing the pdf values for these grid points with 0, and adding P to the distribution at the first grid point above subsistence.

For adjustment 2, in addition, a "richness line" R is introduced. Starting at the upper end of each group, move everyone above the richness line (the total mass of people in the group with income above R) down to the first grid point below R. Then decrease R until this procedure makes the mean of the group equal to the pre-adjustment mean.

Finally, all the group distributions are summed into a population distribution. Then, defining all grid points as discrete groups (ie 50 000 groups), (7) is used to calculate the overall Gini coefficient.

Adjustment 3

The log-normal distribution is now calculated on $X - y_{min}$ instead of on X, for each group (y_{min} is found in the same way as for the previous adjustments). Then, the complete distributions are right-shifted by y_{min} again, before they are added. Then, Gini coefficients can be computed on the grid points in the same manner as for adjustments 1 and 2.

Benchmark

An unadjusted Gini coefficient is also calculated by the numerical method. The largest deviations on the unadjusted Gini comparied to coefficients calculated by Equation 5 are .09 Gini points (.0009) for New Spain and .01 Gini points (.0001) for Chile; this verifies that the numerical procedure is sufficiently accurate to compare the benchmark to the adjusted values.

C.4 Gini and number of groups relationship

A regression of Gini coefficients on the number of groups, for $c = 1$, is shown in Table C.2. The point estimate is very close to zero, and not significant. Brazil (with $N = 375$) is an outlier in terms of number of groups and was not included in the regression shown here. Including Brazil in the regressions does not change the sign or significance level of the coefficients.

	Dependent variable: Gini coefficient						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	0.450***	0.461***	0.464***	0.540***	0.632***	0.743***	0.624***
	(0.025)	(0.025)	(0.024)	(0.018)	(0.013)	(0.008)	(0.026)
Number of groups	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.002	0.000	0.002	0.015	0.023	0.029	0.005
N	24	24	24	24	24	24	24

Table C.2: Lack of correlation between Gini coefficient and number of groups.