COMPARING ANCIENT INEQUALITIES:
THE CHALLENGES OF COMPARABILITY, BIAS AND PRECISION

Fochesato, M., Bogaard, A. and Bowles, S.

The growing literature on ancient inequalities incorporates estimates of wealth distribution from widely diverging periods, cultural contexts and data types (e.g. (Morris, 2010, Scheidel, 2017)). Non-documentary archaeological sources of evidence are the only means of charting wealth distribution for vast stretches of the human past, and different approaches (e.g. grave goods, house sizes etc.) have emerged (Windler,Thiele and Müller, 2013, Kohler and Higgins, 2016, Kohler,Smith,Bogaard et al., 2017, Kohler and Smith (eds.), 2018, Porčić, 2018). Given the unique potential of archaeological evidence to investigate patterning in inequality over the very long term, there is a clear need for robust measures of inequality that are comparable across differing time periods, cultures, technologies and political systems. Enhancing comparability is also important for interpreting local estimates in a wider context.

A natural source of methods for addressing these challenges is economics. Simon Kuznets pioneered the comparative and historical study of the total output of an economy measured by gross domestic product (Kuznets, 1965). More recently Anthony Atkinson, Thomas Piketty and their colleagues have used measures of the share of total income received by the richest members of a society to study economic inequality since the early 20th century (Atkinson and Piketty, 2007).

Notwithstanding the limits of simple measures of complex and multidimensional phenomena, these measures have provided indispensable lenses for the comparative study of modern economies. But their deployment in archaeological studies raises a number of challenges. First, much of the above work is based on a complete enumeration (using census or tax authority data) of the relevant populations, while archaeological studies necessarily rely on samples of data (often quite small and with unknown statistical properties). Second, most economic measures use market prices as a common indicator of value for adding up the heterogeneous elements making up total output or a living standard, while in most archaeological applications, no similar method of aggregation is feasible. Finally, the set of societies that economists typically study is substantially more homogeneous in culture, social structure, and technology than those studied by archaeologists (they typically are similar in having censuses and taxation, for example).

Here we propose methods for measurement of wealth inequality using archaeological data; and we present our resulting set of estimates for 150 Eurasian, North and Central American site-phases, ranging in age from 23,000 years ago to the 18th century CE, building on recently assembled datasets for quantifying ancient wealth inequality. We study inequality in wealth – a stock of assets that yields a flow of income or other valued services over time – rather than income (or some other measure of living standards) because archaeological methods provide a much better basis for estimating wealth than income.

Some of the dimensions along which we measure inequality are best conceived of as individual attributes: that is, something that people simply *have* more or less of, like height. But other dimensions are best conceived of as an aspect of relationships between people, measured by differences in some attribute. Economic inequalities in wealth among households are in this latter class.

Our measure of wealth inequality, the Gini coefficient, ranges from zero (complete equality) to one (one person has all of the wealth) and is defined as one half of the mean of the differences among all pairs of households in the population, divided by the mean wealth. Because they are sometimes available in written tax or probate records, partial measures of inequality such as the fraction of all wealth owned or income received by some small fraction of the population are widely used. In contrast, the Gini coefficient is a measure of inequality in the *entire* distribution of wealth (see also online supplement).

A major challenge to comparative research on ancient wealth inequality is that the relevant information from various sites and phases is based on different sample sizes and methods, pertains to different indicators of wealth (e.g., dwelling or storage area size or grave goods), may omit measurement of those without wealth (e.g., slaves) and may be from populations of vastly differing size (e.g., a city or small hamlet). Here we develop methods for using such heterogeneous data to produce estimates that are comparable across sites and phases.

Cross-cultural work has an important place in archaeology (Bogucki, 1999, Trigger, 2003), not least in informing long-term perspectives on contemporary dilemmas, growing economic inequality being a prominent example. We aim to show that the methods of comparability we propose here offer superior insight than collation of available Gini coefficients without regard to the very different underlying information on which they are based.

There is a limit to what can be said on the basis of even comparably measured indicators of material wealth inequality. Two societies with equal wealth inequality by our measures may differ substantially in social complexity, political hierarchy, disparities in consumption, economic injustice or other dimensions associated with the term wealth inequality. For example, while evidence is indirect, many of the societies under consideration may have practiced systems of redistribution with the result that inequalities in consumption were less than wealth inequality. Our measures are also not directly informative about political inequality, and how this may differ, for example, between stateless and state-governed populations.

Two identical Gini coefficients may even be associated with radically different distributions of wealth, for example one in which inequality arises from a small concentration of entirely propertyless households at the bottom in an otherwise relatively equal population and the other with a few exceptionally rich households in a population of small landowners (e.g., see online supplement). Thus, the Gini coefficient can sometimes mean very different things on the dimensions that are not directly measured.

1 The dataset

The archaeological data incorporate a series of regional datasets assembled by other scholars, plus additional sites with accessible data of particular interest for questions of inequality. All of the sites we include are shown in Figure 7 and listed in the online supplement. Geographically they are distributed across Eurasia, North and Mesoamerica; chronologically they range from a single observation at 23,000-year-old Ohalo II in the southern Levant (Nadel, 2003) to 18th-century AD communities of the Pacific Northwest (Schulting, 1995, Prentiss,Foor and Murphy, 2018). Where possible we distinguish chronological phases within site sequences in order to avoid combining distinct periods of occupation.

Two of the archaeological datasets we include, for the Columbia Plateau (Schulting, 1995), and Hohokam (McGuire, 1992), offer large multi-site data on grave goods in individual burials. We use these datasets, chosen for their size and disparate cultural contexts, to develop some of our adjustments to wealth distribution estimates below. We also bring in historical datasets – on land ownership in a large agricultural population in 17^{th}- 18^{th} century Germany, at Krummhörn (Willführ and Störmer, 2015), on the distribution of household wealth in 1427 Florence, Italy (Herlihy and Klapisch-Zuber, 1985) and on regional inequality from medieval Finland (Nummela, 2011) – to develop methods for refining wealth distribution estimates.

Our unit of analysis for assessing wealth distribution is the household. We define 'households' as co-residential groups occupying modular architectural units with standardized features that suggest redundancy of domestic functions among units. Multiple units may cooperate as a larger household group, but the widespread archaeological observation of standardized domestic units suggests that they often acted as fundamental social agents. Moreover, though wealth may sometimes be shared across households, systematic wealth sharing takes place within the house; the house can be defined as a physical (and metaphorical) unit for storing and sharing wealth (Gudeman and Rivera, 1990).

2 From individual to household inequality in grave goods

Because household membership typically cannot be identified from burial remains, we calculate a between-household inequality measure. From the burial sites of the Columbia Plateau (Schulting, 1995) and the Hohokam culture (McGuire, 1992) we identify those with the greatest number of gender-identified observations. For these four sites we first computed the Gini coefficient of individual wealth (among only the individuals whose gender is identified). Second, we estimated the Gini of couples' wealth, where hypothetical couples were created assuming perfect wealth assortment, i.e. richest females are matched with richest males and poorest females with poorest males. Couples' wealth is then the sum of the wealth of the individuals matched. Third, we computed and averaged ten Gini coefficients on couples' wealth with couples generated assuming random assortment, i.e. males and females were randomly matched irrespective of wealth. Table 1 in the online supplement presents the results. Fourth, our estimate of between-household inequality is the average of the results of the two methods – perfect assortment and the absence of assortment. The ratio of the Gini coefficients estimated in this way for our hypothetical couples to the one estimated from individual data is 0.90 and 0.92 for, respectively, the Columbia Plateau and Hohokam. In subsequent adjustments we use the mean value 0.91.

Moreover, a robustness test for our method is possible using a small dataset (based on ethnographic observations) for which we have the wealth of males and females in actual couples, so we can check if the estimated Gini coefficient for hypothetical couples obtained through our method is close to the Gini coefficient computed on the observed wealth of true couples. In the online supplement we find that the two estimates are similar. The fact that the three estimates, two archaeological and one ethnographic, are very similar motivates our application of this method to the other archaeological cases in our dataset.

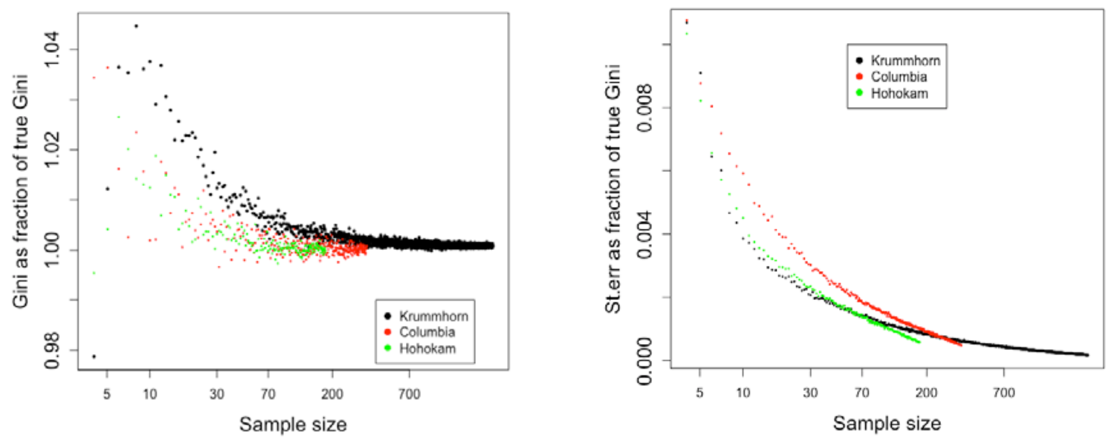
3 Sample size and the Gini coefficient

Because the fundamental data underlying the Gini coefficient are wealth differences among pairs of households, in principle it can be calculated for a population (or from a sample) as small as two households. Here we explore the statistical properties – bias and imprecision – of estimates of a Gini coefficient for wealth based on small samples of m observations from a (sometimes unknown) number M in the total population. The question of interest is: if we recover data on m individuals in an estimated total population of M individuals, what is the nature and magnitude of the biases affecting Gini estimates, and how do larger sample sizes attenuate these biases and the imprecision of the estimates?

Using a dataset with observations of individual wealth ownership, we assume that the total number of observations, M, is the total population and the Gini coefficient computed on M is the true Gini coefficient. We then hypothetically restrict the dataset to a number of individuals $m<M$, and then estimate the Gini coefficient that would have resulted. To do this, we first set $m=2$, then randomly select a pair from the dataset and, on this basis, we calculate a Gini coefficient. We conduct this process of sampling with replacement for $m = 2$ a thousand times, producing a mean estimate and its standard error. We do the same for all values of $m < M$.

We implement the algorithm on three large datasets from both archaeological and historical records: the whole burial wealth dataset from the Columbia Plateau (Schulting, 1995), and from Hohokam society at la Ciudad (McGuire, 1992) and the records of land ownership in a large agricultural population in 17^{th}- 18^{th} century Germany, at Krummhörn (Willführ and Störmer, 2015).

The results are shown in Figure 1. Panel (a) shows that bias is substantial when the sample is very small and it quickly approaches zero as the sample size m increases. Neither skewness nor total population size (M) appear to affect the extent of bias in this case. Panel (b) of Figure 1 shows that the standard errors of the estimated Gini coefficients (as a fraction of the true Gini coefficient) are strikingly small, even for samples of modest size. That the relationships in Figure 1 are very similar despite being based on very different cultural contexts suggests that our statistical method has wider applicability across our heterogeneous dataset.



(a) Estimated Gini as a fraction of the true Gini (b) Standard errors as a fraction of true Gini

Figure 1. **Sample bias and standard error of the Gini coefficients in three large datasets.** The (a) and (b) panels, respectively, show the estimated Gini as a fraction of the true Gini and the standard errors of the estimated Gini coefficients as a fraction of the true Gini, for Columbia Plateau (red dots), Krummhörn (black dots), and Hohokam (green dots). The x-axis is ratio scaled.

Adjusting Gini estimates for sample bias. To adjust downwards the Gini coefficients that are upward-biased because they are based on small samples, we use the Columbia Plateau dataset shown in Figure 1 and estimate a non-parametric relationship between the bias and the natural logarithm of the sample size. (The adjustment would not be appreciably different had we used the other datasets, given the very similar relationships of sample size and bias shown in the above figures.) We summarize the relationship between the two variables using a local polynomial regression (Figure 2). For each Gini coefficient in our dataset, we use the number of observations in our data to predict the sample bias so as to estimate the size of the Gini that would have been estimated had the entire population been observed. This procedure does not require us to know the true population size, since the extent of bias becomes very small for samples of modest size, irrespective of M.

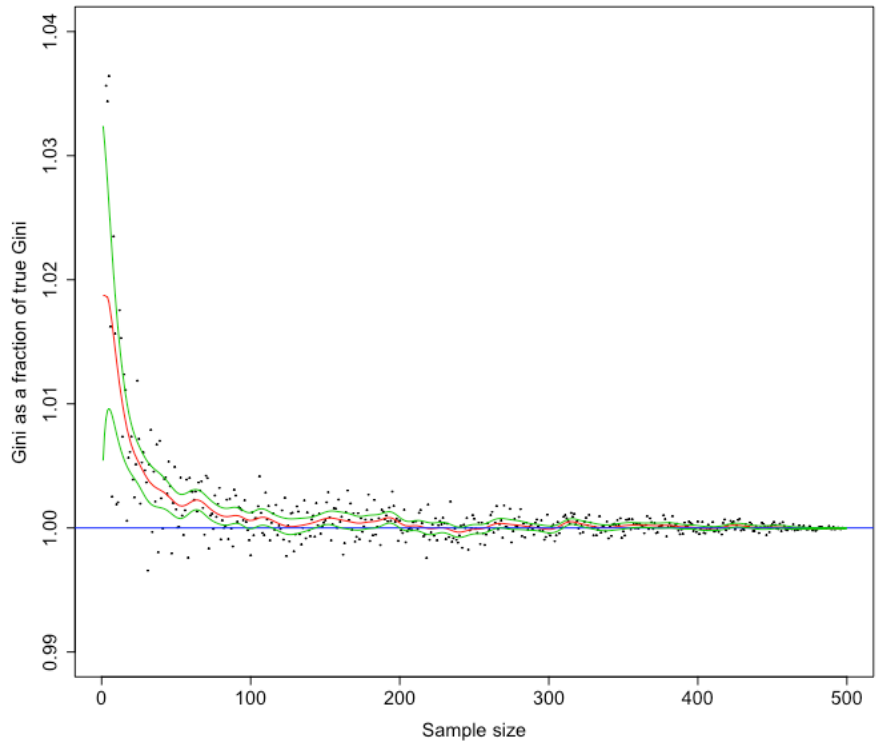


Figure 2. Non-parametric regression for sample bias. The estimated Kernel regression (red line) of the ratio of the estimated Gini to the true population Gini (black dots) on the population level. The bandwidth is equal to 5. Shown are also the confidence intervals, derived as plus and minus twice the standard error.

4 Accounting for those without wealth

Due to data limitations, estimates of Gini coefficients often omit relevant zeros, meaning those with none of the attribute being measured. Inequality of land ownership, for example, is often mis-measured by a Gini coefficient on the holdings of land owners, excluding the landless. For example, estimates of inequality of grave goods in southern Mesopotamia and Roman Italy (Stone, 2018) include those without wealth among the free population, but exclude the slaves who lived and worked in the urban centers. The result of these exclusions is to understate substantially the degree of wealth inequality in a population.

To include these missing non-owners ('zeros') we need two pieces of information. The first is how numerous these missing zeros are, which we estimate from historical sources about the population in question. The second is an estimate of the effect of excluding zeros on the Gini coefficient, which we obtain by manipulating a mathematical expression for the Gini coefficient in a class divided economy. It can be shown (Bowles and Carlin, 2018) that an approximation of the true Gini coefficient (G) based on an observed estimate (G') calculated from a dataset with a fraction u of the total population without wealth not used in the estimate (so u is the fraction of 'missing zeros' in the total population) is

$$G = G' + u - uG' \tag{1}$$

We use this expression, along with an estimate of u to calculate the true Gini.

We check the validity of this method by the same methods already introduced to estimate the effects of small sample size. We study populations for which we know the entire distribution and from which we can hypothetically remove the zeros and then compare the true estimates with our estimates based the hypothetical absence of data on the zeros using equation 1.

We use 32 distributions of different forms of wealth: grave goods of 23 burial sites from the Columbia Plateau (Schulting, 1995), and four from the Hohokam (McGuire, 1992), the distribution of household wealth in 1427 Florence (Herlihy and Klapisch-Zuber, 1985) and four cross sections of land ownership in Krummhörn, Germany, three centuries ago (Willführ and Störmer, 2015). Table 3 in the online supplement shows the data used for the analysis. To check our method we estimate, for each population, the Gini coefficient of the whole distribution using the known fraction of zeros in column 3 and the Gini coefficient estimated on the hypothetical population for which the zeros have been removed.

The mean absolute error between the estimated and the true Gini on the total population as a fraction of the mean true Gini is 0.012, and the correlation coefficient between the two sets of values is 0.99 ($p<0.001$). These results suggest that our method is reasonably precise, and it provides the basis for our upwards adjustment of the Gini coefficients with missing zeros. As a further check, we use a least squares regression to predict the true (entire population) Gini coefficients using the values of g', u and ug' derived by hypothetically removing zeros from these populations. As predicted from equation (1), the estimated regression coefficients are almost exactly one (for the first two) and minus one for the product. We describe how we estimated the numbers of those excluded – landless slaves in the case of southern Mesopotamia and Roman Italy – in the online supplement.

5 Comparability among different asset types

Some asset types tend to be more equally distributed than others; in modern societies, for example, housing is much more equally distributed than ownership of companies. If we want to compare the inequality of household wealth in different societies, we first need to assess how the distribution of the asset that is available to measure inequality in a specific society compares to the distribution of the other forms of wealth constituting household wealth.

We identify the following four measures on which material wealth inequality has been estimated: land, house storage space, house living space and grave goods. Determining what counts as wealth is a critical issue. In agricultural societies, even in the small-scale labour-intensive ones included in our dataset, at least informal property rights in land are likely to exist and the main source of well-being production for a household was the land it cultivated. For this reason, when both living and storage spaces are clearly identified, we use only storage area (indicative of land inequality) as a proxy for household wealth. However, in many archaeological sites living and storage areas have not been distinguished. In these cases, we use the total house area as proxy of household wealth. For some settled agrarian societies, household wealth inequality is measured directly through land inequality. In these cases, given the primary function of land for the production of household well-being, we take it as the measure of household wealth.

We consider grave goods not as a form of wealth but as a costly signal, conveying information about the status and wealth of the deceased. While the goods that make up burial assemblages may sometimes be objects making up an individual's or group's wealth – tools, weapons, animals and valued household objects – grave goods may also be entirely symbolic and non-utilitarian. What matters for our purposes is that, whatever their form, burial goods are an indicator of the wealth of the household of the deceased because the household must forgo some of its wealth to provide grave goods for the burial assemblage.

There are both reasons and evidence to support the hypothesis that grave goods are more unequally distributed than forms of wealth such as indicated by dwelling size (see online supplement and Peterson and Drennan (2018)). In southern Mesopotamia, during the Neo-Babylonian period, for example, Gini coefficients for house area (living and storage space) and grave goods were respectively 0.621, and 0.878 (corrected by sample bias, couples and missing zeros as explained in the previous sections.)

We reconcile these two wealth types – house area and grave goods – using archaeological datasets for which inequality is measured in the same society and in the same time period for both asset types. For the cases in which the two measures are available (Table 5, online supplement), the inequality in house area (indicative in this case of household wealth) is on average 71.9 percent of the inequality in grave goods, with a remarkably small standard deviation of 4.7%. We infer from this that, were household wealth data available for those societies on which we have data on grave goods only, the inequality in the latter would be approximately three quarters the inequality in the grave goods. Therefore, a Gini coefficient measured on grave good inequality alone will be reduced by 28.1 per cent of its value to make it comparable to inequality in household wealth.

6 Scale effects: Comparability across different population sizes

Suppose that we have data on a single 'village' but for reasons of scale comparability with our other estimates, we would like to estimate the degree of inequality in the 'district' of which that village is a part, along with the other villages making up the district, but on which we do not have data. Taking wealth inequality in the 'village' as an estimate of wealth inequality in the entire 'district' will not produce comparable estimates because we expect that larger populations will be more heterogeneous geographically, demographically and even institutionally and culturally and hence may exhibit greater levels of wealth inequality. We can see that this is the case in our measures of inequality of grave goods on the Columbia Plateau. For example, the Gini coefficient for the entire population in the late prehistoric phase is 0.647, while the average of the Gini for the six burial sites in the same phase is 0.573.

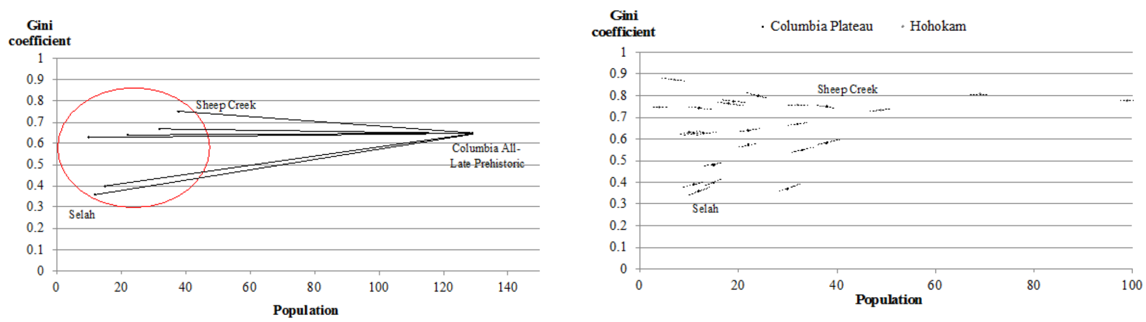
To achieve comparability of scale, we use an estimate of the population size effect that provides the Gini coefficient that hypothetically would obtain were the population in question of some benchmark size. The method is based on comparisons of Gini coefficients for lower-level population entities and the larger entities that they make up. We call this the 'nested method'. The advantage of the nested method is that we are able to estimate the size effect for population groups that are likely to be similar in most respects other than size, because the larger unit is composed of the smaller units. It provides a far more accurate estimate of the pure scale effect than is possible using non-nested data, namely by comparing Gini coefficients across populations of differing sizes.

Three datasets allow us to estimate the difference between the Gini for the constituent lower-level units and the Gini for the higher-level unit that is the composite of all of these: the Columbia Plateau dataset (Schulting, 1995), the Hohokam dataset (McGuire, 1992) and a dataset for 1571 Finland (Nummela, 2011), the latter representing a pre-industrial dataset of wealth distribution with complete geographical coverage (each upper administrative unit is the composite of all the lower ones).

In the online supplement we use the Columbia Plateau dataset to explain how the scale effect is computed. The scale effect is illustrated in Figure 3(a) by the slope of the line connecting the Gini at the site level (the smaller entity) to the Gini of the entire population made up of these six sites. In Figure 3(b) we compare the scale effects for all of the phased Columbia Plateau sites with those of the Hohokam sites, showing that as the population of the lower level entity increases, the scale effect (represented by the slopes of each point in the figure) decreases.

To ensure coverage over the entire range of relevant population sizes, we merge the three datasets with the scale effects measured at lower levels with population smaller than 250 households (Columbia Plateau, Hohokam and late medieval Finland) and plot in Figure 4 the relationship between the scale effect (multiplied by 1000) and the population of the lower units at which it was computed. We observe that, for very small populations, the scale effect is substantial and, as population increases, a very sharp decline in the scale effect occurs; beyond a population sizes equal to about 50 households, the scale effect is close to zero. We develop a statistical summary of these data that will allow us to scale-adjust the Gini coefficient for any population size.

The estimated scale effects at each population level are used to derive a function reflecting what we call ‘estimated pure scale effect’ showing how the Gini varies over the range of population sizes for reasons of scale alone. For the purposes of our adjustment what matters is its slope, not the value of the Gini for any population level. We arbitrarily chose as a starting value of our function a Gini equal to 0.5 at the benchmark population level equal to 50 households. From this arbitrary benchmark and the estimated slopes in Figure 3(b), we construct the estimated pure scale effect curve in Figure 5.



(a) (b)

Figure 3. Examples of scale effect for Columbia Plateau and Hohokam. Panel (a) shows, the Gini coefficients and population size for each site of Columbia Plateau in the late prehistoric phase (dots inside the red circle) and the Gini coefficient and population size at the regional level in the same phase. Panel (b) The scale effects for each burial site at the Columbia Plateau excavation (black dots) and at the Hohokam excavation (grey dots). Here the scale effects are the slopes of the line segments at each point, illustrated in **Figure 3**(a) by the slope of the line from the lower level entity to Columbia.

Figure 5 shows how the function is used to size-adjust the observed Gini coefficient for Neolithic Vahingen, Germany (after the correction by sample bias, 0.189) estimated from a population of 11 households. We let g^i be the observed Gini, $g_i(50)$ the size corrected Gini, $g(11)$ the predicted Gini for population size equal to 11 households and $g(50)$, the predicted Gini for population size equal to 50 households. The figure also shows that the resulting scale correction, equal to the difference between $g_i(50)$ and g_i is 0.007.

We can assess the accuracy of our method by the following thought experiment. Suppose there are M a large number of lower level entities "villages" that make up a district, and we have evidence on $m < M$ of these, i.e. some but not all of the villages. How accurate a prediction of the inequalities at the district level will we produce using our estimated pure scale effect function estimated from our three data sets and shown in Figure 5? Using the largest set of lower level entities from the Columbia Plateau data set (M=10) to predict the district level inequality, we find that when we use all 10 of the data points the error is 0.04, and with m=4 the error is 0.06.

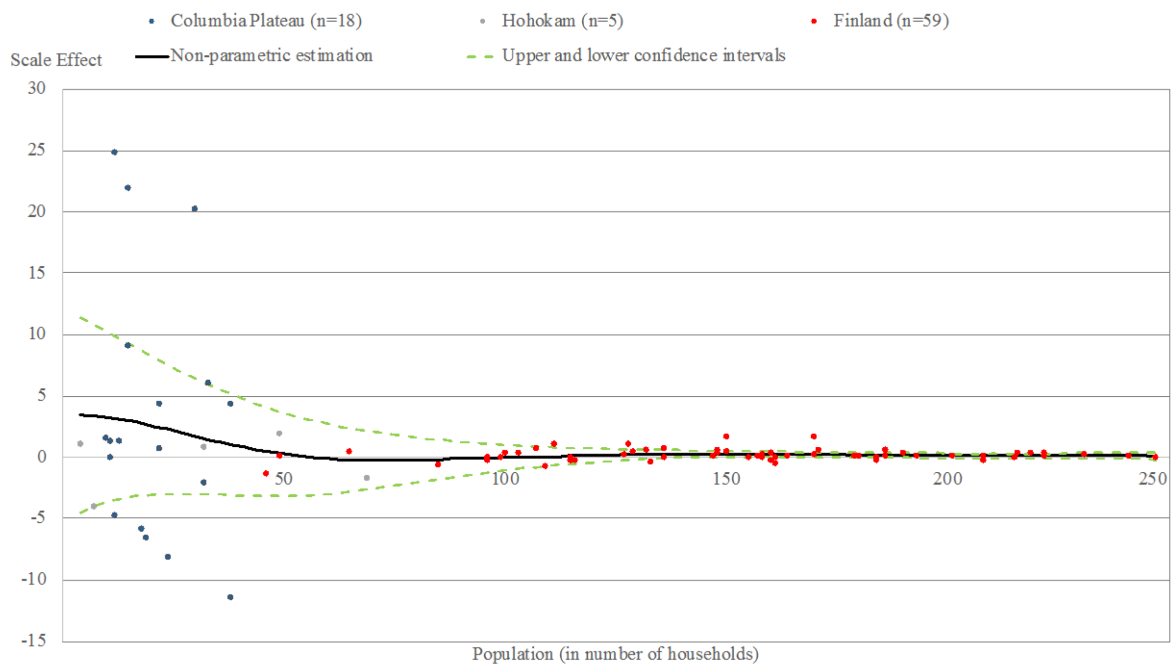


Figure 4 Scale effect estimated from Columbia Plateau, Hohokam and Finland datasets and kernel regression. The scale effects estimated from the Columbia Plateau (blue dots, n=18), Hohokam (gray dots, n=5) and Finland (red dots, n=59) datasets and the non-parametric relationship between the scale effect and the population level (black line). Shown are also the confidence intervals, derived as plus and minus twice the standard error.

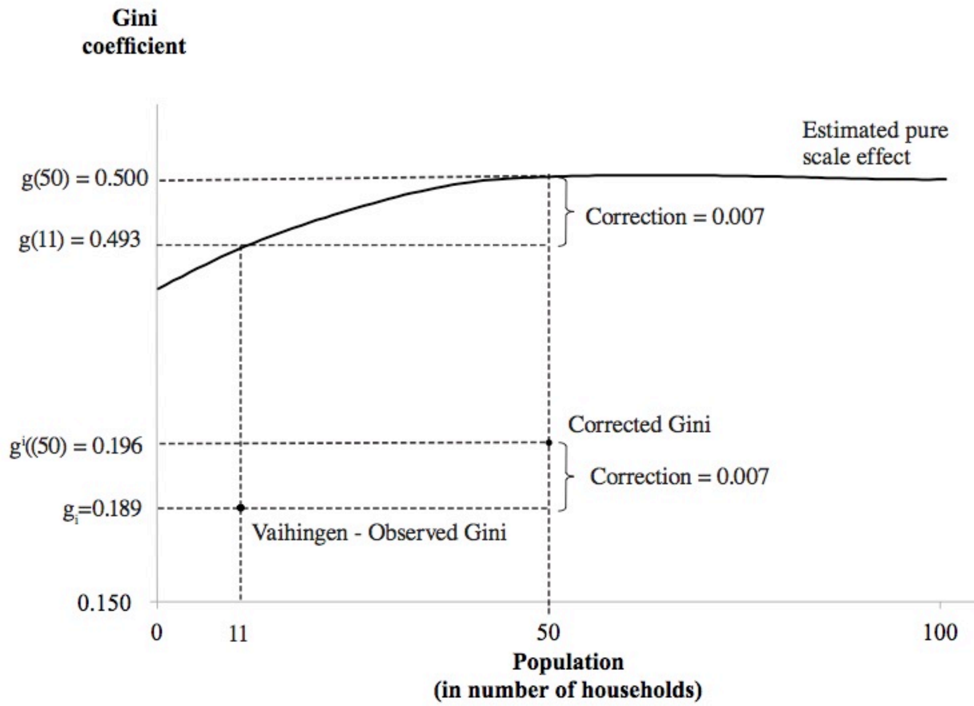


Figure 5. Scale adjustment to a common population level. An example of scale adjustment when the actual population is smaller than the benchmark population level (50 households). The example is the adjustment of the Gini coefficient for Neolithic Vaihingen (Germany, 6^{th} millennium BC), where the original Gini was estimated on 11 households. The adjustment works in the same way when the raw Gini was estimated on a population larger than the baseline.

7 Results

Figure 6 shows the relationship between the Gini coefficients entirely unadjusted for comparability, and the fully adjusted estimates. Where adjustments have been limited to taking account of population and sample size, the final estimates are quite similar to the unadjusted ones, because except for very small populations and sample sizes the biases that we have estimated are quite modest. By contrast, the adjustments for estimates based on grave goods (downwards, that is to the right of the 45 degree line in the figure), or for excluded slaves and other households without property, the adjustment (upwards, that is to the left of the 45 degree line) are substantial. The average absolute value of the adjustment is 15 percent of the value of the raw Gini coefficient, suggesting that unadjusted measures are in general quite unreliable.

The 150 adjusted Gini coefficients included in our dataset show a wide heterogeneity in the level of inequality from archaeological sources (Figure 7) ranging from about 0.1 to almost 0.8 Elsewhere we use these data to provide an interpretation of the increase in wealth inequality in western Eurasia up to the early first millennium AD (Bogaard,Fochesato and Bowles, Submitted).

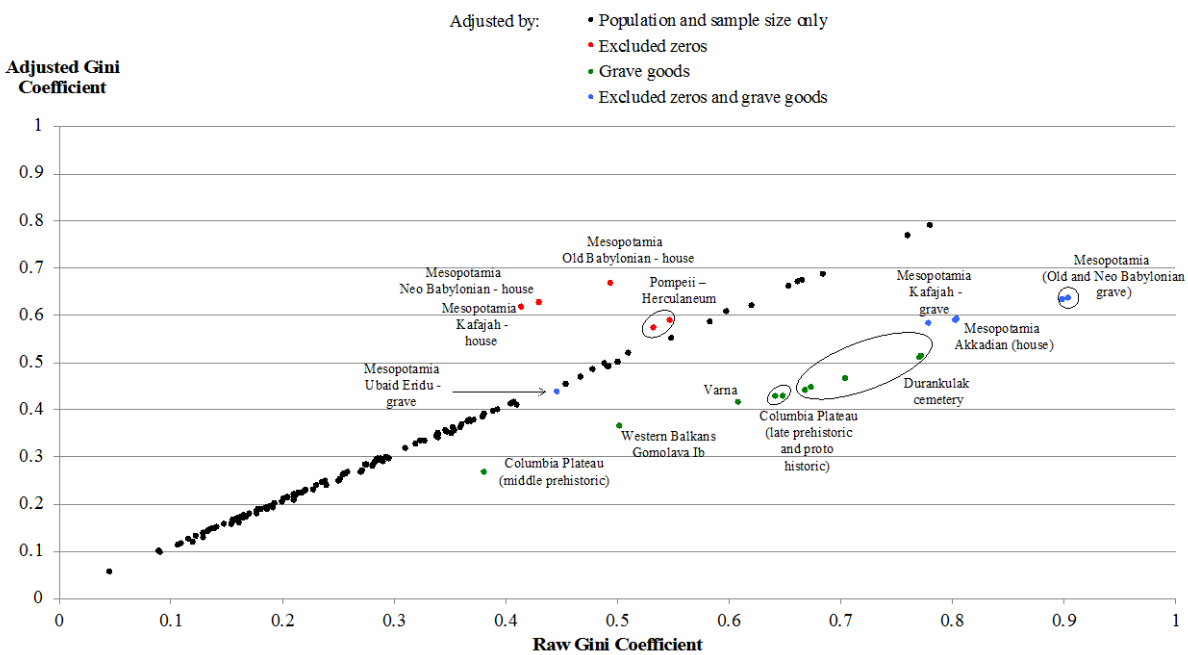


Figure 6 Comparing the raw and adjusted Gini coefficients. The relationship, for each case, between the raw Gini coefficient (horizontal axis) and the Gini coefficient after all the adjustments (vertical axis). The black dots are the estimates adjusted by population and sample size only, while red, green, and blue dots are the Gini that have also been adjusted by, respectively, the missing fraction of zeros, the grave goods, or both adjustments. Source: see online supplement.

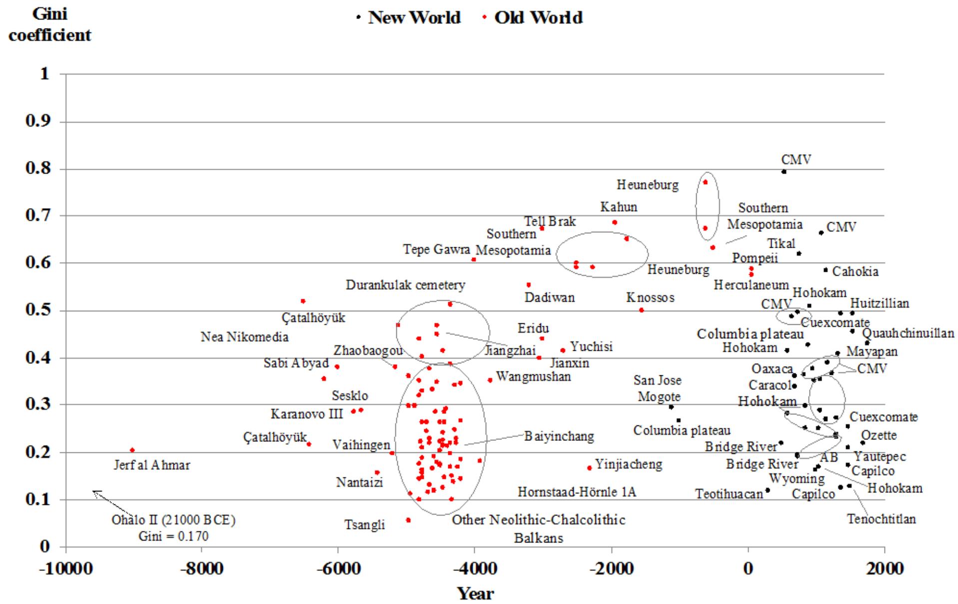


Figure 7 Wealth inequality from archaeological data. The Gini coefficients in the 150 societies included in the dataset, after the adjustments described in the previous sections (CMV = Central Mesa Verde; AB = American Bottom). Sources: see online supplement.

8 Discussion

We have sought to measure wealth inequality in ways that make it substantially comparable across diverse economies, cultures, and time periods. In some respects, the results are promising. Our demonstration that relatively small samples from much larger populations yield quite accurate and precise estimates of the Gini coefficient is encouraging, especially for archaeologists and others unavoidably constrained by the feasibility of generating more inclusive samples. But the underlying assumption that what is "found" and what is "missing" is random rather than systematic is likely to be violated in practice. Where there is a clear bias towards e.g. excavation of larger houses, as at Knossos (see online supplement), or the exclusion of slaves and other households without property, our adjustments have addressed the bias as adequately as current data allow.

Our methods of approximation of the total population Gini coefficients from estimates where those without wealth are missing also appear to be surprisingly accurate. Similarly, our estimates of wealth inequality among couples hypothetically constructed from individual-based data are consistent across two archaeological sites and replicated almost exactly using a sample (albeit small) for which complete data are available.

But in another respect, our results highlight uncertainties. We lack prices or similar valuations as a method of establishing comparability across types of assets (housing versus grave goods, for example) or even among differing indicators of similar assets (the heterogeneous objects making up our measures of grave goods). In the latter case we used systems of relative grave good values adopted by the archaeologists who initially described the data. In the former case – comparing distinct asset types – we used a relatively small number of datasets in which more than a single dimension of wealth has been measured. The fact that the estimates on which these conversions are made (grave good inequality to house inequality) are very similar across datasets and that grave good inequality is very highly correlated with house size inequality (suggesting that the former is informative about the latter) is reassuring.

Moreover, our measures are necessarily incomplete. We have not, for example, measured the human wealth of slave owners, which in some societies in our sample (e.g., southern Mesopotamia) would constitute a considerable fraction of total wealth. We have also not attempted to incorporate systematic measurement of livestock wealth.

These uncertainties and gaps, while substantial, should not be exaggerated. On the basis of currently available data, we cannot think of any plausible adjustment in the data that would alter the impression from Figure 7 that Neolithic populations in western Eurasia tended to experience considerably less wealth inequality than many of the later cases, and that post-Neolithic wealth inequality in Eurasia tended to be higher than inequality in the western hemisphere. We have thus shown that more systematically and comparably measured indicators of wealth disparities do not overturn and indeed reinforce the primary finding of Kohler,Smith,Bogaard et al. (2017).

Acknowledgements

Thanks to Paul Halstead, Rick Schulting, Todd Whitelaw, Chiaki Moriguchi, seminar participants at the Hitotsubashi University Institute of Economic Research and participants of the 2017 Oxford short course on long-term inequality and two anonymous reviewers for insightful comments; Marko Porčić for his unpublished database of Balkan Neolithic-Chalcolithic house sizes; Kai Willführ and Eckart Voland for providing the updated Krummhörn dataset; and the Dynamics of Wealth Inequality Project, Santa Fe Institute, for support and hospitality.

ATKINSON, A. & PIKETTY, T. (eds.) 2007. *Top Incomes Over the Twentieth Century: A contrast between contenintal Europe and English-speaking countries,* Oxford: Oxford University Press.

BOGAARD, A., FOCHESATO, M. & BOWLES, S. Submitted. The farming-inequality nexus: new methods and archaeological evidence. *Antiquity*.

BOGUCKI, P. 1999. *The Origins of Human Society*. Oxford: Blackwell.

BOWLES, S. & CARLIN, W. 2018. Inequality as experienced difference: A reformulation of the Gini coefficient. *London, Centre for Economic Policy Research*.

GUDEMAN, S. & RIVERA, A. 1990. Conversations in Colombia: The Domestic Economy in Life and Text. Cambridge: Cambridge University Press.

HERLIHY, D. & KLAPISCH-ZUBER, C. 1985. *Tuscans and their families: a study of the Florentine catasto of 1427*. New Haven: Yale University Press.

KOHLER, T.A. & HIGGINS, R. 2016. Quantifying household inequality in Early Pueblo villages. *Current Anthropology* 57.

KOHLER, T.A. & SMITH (EDS.), M. 2018. *Ten Thousand Years of Inequality. The archaelogy of wealth differences.*: Amerind Foundation.

KOHLER, T.A., SMITH, M.E., BOGAARD, A., FEINMAN, G.M., PETERSON, C.E., BETZENHAUSER, A., PAILES, M., STONE, E.C., MARIE PRENTISS, A., DENNEHY, T.J., ELLYSON, L.J., NICHOLAS, L.M., FAULSEIT, R.K., STYRING, A., WHITLAM, J., FOCHESATO, M., FOOR, T.A. & BOWLES, S. 2017. Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. *Nature* 551: 619.

KUZNETS, S. 1965. Economic growth and income inequality. *American Economic Review* 65: 1-28.

MCGUIRE, R.H. 1992. *Death, society, and ideology in a Hohokam community*. Boulder: Westview Press.

MORRIS, I. 2010. *Why the West Rules - for Now. The Patterns of History, and What They Reveal About the Future*. Farrar, Straus and Giroux.

NADEL, D. 2003. The Ohalo II brush huts and the dwelling structures of the Natufian and PPNA sites in the Jordan Valley. *Archaeology, Ethnology & Anthropology of Eurasia* 1: 34-48.

NUMMELA, I. 2011. Regional inequality, in M.H. LAMBERG, M.; J. HAIKARI, J. (ed.) *Physical and Cultural Space in Pre-Industrial Europe. Methodological approaches to spatiality.* Nordic Academic Press.

PETERSON, C.E. & DRENNAN, R.D. 2018. Letting the Gini Out of the Bottle: Measuring Inequality Archaeologically, in T.A. KOHLER & M.E. SMITH (eds.) *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences.* Amerind Foundation.

PORČIĆ, M. 2018. Evaluating social complexity and inequality in the Balkans between 6500 and 4200 BC. *Journal of Archaeological Research*.

PRENTISS, A.M., FOOR, T.A. & MURPHY, M.-M. 2018. Testing Hypotheses About Emergent Inequality (Using Gini Coefficients) in a Complex Fisher-Forager Society at the Bridge River Site, British Columbia, in T.A. KOHLER & M.E. SMITH (eds.) *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences.* Amerind Foundation.

SCHEIDEL, W. 2017. *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-first Century*. Princeton: Princeton University Press.

SCHULTING, R.J. 1995. *Mortuary Variability and Status Differentiation on the Columbia-Fraser Plateau*. Burnaby, B.C.: Archaeology Press, Simon Fraser University.

STONE, E.M. 2018. The Trajectory of Social Inequality in Ancient Mesopotamia, in T.A. KOHLER & M.E. SMITH (eds.) *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences.* Amerind Foundation.

TRIGGER, B.G. 2003. *Understanding Early Civilizations*. Cambridge: Cambridge University Press.

WILLFÜHR, K.P. & STÖRMER, C. 2015. Social Strata Differentials in Reproductive behavior among Agricultural Families in the Krummhörn Region (East Frisia, 1720-1874). *Historical Life Course Studies* 2: 58-85.

WINDLER, A., THIELE, R. & MÜLLER, J. 2013. Increasing inequality in Chalcolithic Southeast Europe: the case of Durankulak. *Journal of Archaeological Science* 40: 204-210.