PNAS RESEARCH ARTICLE | ECONOMIC SCIENCES OPEN ACCESS

Accounting for uncertainty and bias in archaeological and historical evidence on wealth inequality

Mattia Fochesato[a,b,1] and Samuel Bowles[c,d]

Affiliations are included on p. 11.

Edited by Michael E. Smith, Arizona State University, Tempe, AZ; received June 5, 2025; accepted April 17, 2026, by Editorial Board Member Christopher B. Barrett

The archaeological and historical record allows us to study the patterns of economic inequalities associated with sharply contrasting institutions such as communal property in societies without states, private property among small holders, and slavery, in economies relying on radically different technologies, based for example on human energy alone, other animal power, and carbon-based power. But the price of using prehistoric and historical data to expand the range of institutions and technologies under study is a substantial and typically unknown level of uncertainty and unrepresentativeness in the resulting estimates. We provide methods and code that quantify the degree of uncertainty, an approach that we term BRIDGE (Bayesian-Resampling and Informed Priors with Data-driven Gini Estimation). We transform raw wealth data from 431 sites and dates into probability distributions representing likely levels of inequality that (insofar as possible) are representative of the underlying population (e.g., rectifying biases arising from small or nonrandom samples and the frequent absence of data on those entirely without wealth) and are comparable across differing asset types (e.g., burial goods, dwelling area, storage area, land cultivated), ownership units (individual or household), and scale (from villages to nations). These distributions are based on the raw data along with independent information that allows us to infer biases and levels of uncertainty. We also account for the propagation of uncertainty through every stage of the data generation and estimation process. We find that widely used methods (e.g., relying solely on conventional bootstrapping) provide misleading and for the most part substantially underestimated measures of uncertainty.

uncertainty | wealth inequality | ancient economies | Gini coefficients | Bayesian statistical methods

Significance

In our era of fundamental changes in technology and institutions along with rising inequality, our assessments of likely future trends and policy options to address inequalities may be enlightened by the historical and prehistoric record of similarly vast changes. But the available data and how it came to be available bears with it the cost of substantial but unknown levels of uncertainty. Going beyond conventional bootstrapping and integrating a suite of statistical tools we provide a comprehensive method and code for estimating levels of bias and uncertainty in the available data on past wealth inequalities and how these uncertainties propagate through every stage of the estimation process. Our estimates of past wealth inequality are probability distributions over likely values.

The many insights provided by recent quantitative studies of economic inequality in the distant past have unavoidably been seen as bracketed by substantial error bars (1–11). What Oskar Morgenstern termed "the fact of error" (12) is endemic to research in this area; but convincing estimates of the associated uncertainty are rarely available.

The point estimates of inequalities in wealth or other economic metrics on which these and related studies are based are often taken as the raw data on which hypotheses can be tested using conventional measures of statistical significance. But these point estimates are themselves each subject to uncertainty that in many cases is substantial. The uncertainties in these data arise in part from the limited number of mostly small-n datasets that are characteristic of archaeological evidence, but substantial uncertainty is also present in estimates based on historical data, especially for sites and dates prior to the introduction of population-wide taxation by modern national states.

In all cases uncertainty is compounded by adjustments to the raw data that are required to make them representative of the underlying populations of interest and comparable across sites and dates. These adjustments are themselves subject to uncertainties beyond the uncertainty of the raw data itself, which then propagate through the estimation process. Ignoring these uncertainties is to claim "specious accuracy" (another Morgenstern expression).

Recognition of the extent of uncertainty surrounding point estimates of ancient wealth inequalities could reasonably lead researchers to eschew quantitative studies on these topics. Or it could motivate the creation of probabilistic measures of wealth inequality based on quantitative estimates of the uncertainties, even if these estimates are themselves highly uncertain.

We here take the second course, implementing a suggestion made by Arthur Bowley and his coauthors over a century ago. He urged the Royal Statistical

Author contributions: M.F. and S.B. designed research; M.F. and S.B. performed research; M.F. analyzed data; and M.F. and S.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.E.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: mattia.fochesato@unibocconi.it.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2514309123/-/DCSupplemental.

Published June 15, 2026.

Society to collect and present data on the distribution of household incomes so as "to tabulate our knowledge or ignorance ... as definitely as possible" thereby allowing "a measurement of precision of the aggregate. ... We are bound to make estimates in all cases, whatever our ignorance may be" and so Bowley explained, "we have endeavoured to assign, in light of all the information available...limits within which it seems highly probable that the true measurement must lie." (13, p. 46–47)

Pioneering works in the measurement of national income by Simon Kuznets and others (14, 15) followed Bowley's advice. But his approach was never widely used by economists (perhaps because of improved measurement of the components of national income). Nor have the methods Bowley advocated been adopted by archaeologists or historians studying past inequalities, notwithstanding the "plea for errors" made by the economic historians Charles Feinstein and Mark Thomas (16).*

Here we provide methods (and code) for estimating probability distributions of likely values of the Gini coefficient of material wealth—a standard measure of inequality (20)—at a particular site and date. We illustrate these methods, drawing on wealth data from throughout the world over the past 12,000 y (21).

An example of one of these distributions—from Akkadia in ancient Mesopotamia—is in Fig. 1. When we refer to "the Gini coefficient for third millennium BCE Akkadia," it is this distribution that we mean. Here and throughout, the distribution-free CI that we report are the least span encompassing 95% of the probability mass of the density function, referred to as CI*'s to distinguish them from conventional (symmetric-around-the mean) CI. Further information on our methods underlying Fig. 1 and the other results reported below appear in *SI Appendix*.†

These distributions differ conceptually from conventional bootstrapped estimates based on repeated sampling with replacement of a given dataset in that we have leveraged additional information on what is known or can be inferred about how the data came to be available and other independent data. Not surprisingly, in addition to our distributions not being Gaussian by construction, they also differ quantitatively, in most cases returning much larger estimates of uncertainty than estimates based on bootstrapping methods.

Our concerns in this paper are entirely statistical, addressing matters of bias and uncertainty rather than substantive questions concerning the Gini coefficient as a measure of inequality, wealth disparity as an indicator of societal inequality, or causal mechanisms likely to account for the patterns of wealth inequality that we observe, topics that we have addressed in related papers (10, 21).

BRIDGE: Bayesian-Resampling and Informed Priors with Data-Driven Gini Estimation

By material wealth (hereinafter, "wealth") we mean alienable assets (e.g., tools, livestock, stored goods, durable household goods, and land) that provide a flow of services (income, reproductive success, or other) contributing to the living standard of a household or other ownership unit, as valued by its members.

To measure wealth inequalities, we would ideally have population-level data from randomly selected sites and dates, measuring a suitably aggregated measure of wealth, or at least a form of wealth that is common across our temporal and geographical scope, held by consistently defined households or other wealth-sharing units. What instead we have at our disposal are measurements of different forms of wealth, some held by individuals, others by households, often from small samples that in many cases are unlikely to be representative of the population of interest, that are distributed across time and space according to the interests and available resources of researchers, and that are from underlying populations vastly differing in scale (and hence in degree of heterogeneity in wealth relevant characteristics). The corrections and adjustments on which our dataset is based are an attempt to estimate both what the researcher would have found if they had been able to collect the ideal data as well as the degree of uncertainty due to the fact that this was not the case.

In many cases, the uncertainty in our estimates arises because the archaeologist, tax collector, or other source of our raw data very often had access to only a subset of evidence, representing a sample of the larger population of interest, whose complete discovery is unlikely or impossible. To estimate the resulting biases and uncertainties we attempt to replicate what might have been the source's search process, sampling from a more complete set of data to estimate what the researcher would have most likely "found" and "not found."

Our methods necessarily differ depending on the extent to which it is likely that the data that have come to light are a random sample of the population of interest. For example, to estimate the uncertainties associated with small sample size per se (rather than the representativeness of the sample) we proceed as follows. We use a large dataset which we consider to be the relevant hypothetical entire population, and then estimate Gini coefficients from arbitrary subsets of these data of differing sizes that could represent the limited information found by the researcher. We use what we call this not found exercise to estimate the uncertainty and bias of estimates for each possible sample size up to a complete enumeration. Then we use the distributions generated for each of the sample sizes in our dataset to estimate the likely bias and degree of uncertainty of our estimates from other datasets where we are limited to small samples without evidence from a larger (hypothetically complete) dataset.

But adjusting for the bias and uncertainties of small random samples is far from sufficient. Suggestive of the nonrandom nature of the search procedures that generated our data is the fact that in a set of ancient Mesopotamian data based on excavations, the frequency of public architecture is twice that of domestic architecture, while in satellite-based data on the same sites, the reverse is true (22).

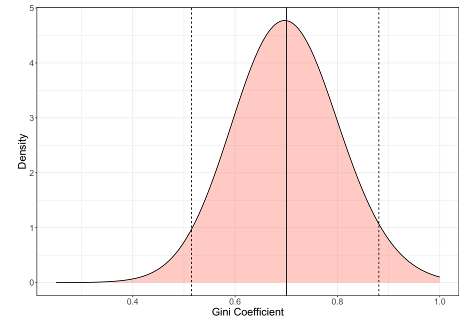


Fig. 1. Probability density function of the adjusted Gini coefficient in Akkadian Mesopotamia (2350–2150 BCE). Solid line: mean (0.700); dashed lines: CI* bounds [0.510, 0.878]. Skewness = 0.089.

*Kuznets followed Bowley's approach in his measurement of national income but his methods were immediately questioned (rightly) by Bowley and Stone (15, 17, 18). Feinstein and Thomas illustrate their method simply by assigning subjective CI to the components of a series on total output. With coauthors we have used the more complete methods described in the pages that follow ref. 19. We are unaware of any other archaeological or historical research using the comprehensive methods that we develop here.

†Full replication code and data are available at https://doi.org/10.3886/E248045V2.

Some of our datasets exhibit glaring evidence that the sample is not representative of the population of interest. A sample entirely of households with substantially above average wealth is probably from a wealthy neighborhood of an unobserved larger settlement, for example. Similarly a sample of excavated house sizes including a palace and the dwellings in a small fraction of the remainder of the occupied area of a settlement cannot alone provide the data for Gini coefficient that is informative about the level of wealth inequality in the settlement as a whole. In these and similar cases we have either excluded the evidence entirely as unrepresentative or used the available evidence to simulate a more complete sample, for example using the available data on commoner housing to simulate a distribution of house sizes in the entire occupied (but not yet excavated) area of the settlement. A related problem arises where the observations are from small settlements that are not representative of the scale of our benchmark population. In these cases we use additional datasets allowing estimates of the degree of bias and uncertainty entailed by sampling from just one or a few of the many subunits making up a settlement of our benchmark scale.

Here we provide an overview and the conceptual framework of our adjustment method. We identify five sources of potential bias and imprecision that may drive a wedge between the observed and true value: sample bias from small datasets, difference in wealth ownership unit (individual- vs. household- level wealth), missing data for households without wealth (or more generally, the less wealthy), the appropriate aggregation of different asset types, and taking account of different population scales. For each of these five sources of bias and uncertainty, we design an adjustment procedure that specifies both the parameter capturing the bias or imprecision, and the distribution of that parameter.

In each case, the choice of method depends on the nature of the adjustment and the availability of additional data that may contribute to an estimate of the likely degree of bias and uncertainty. For the problem of sample bias in small datasets, we derive the parameters for the adjustment and their uncertainty using nonparametric bootstrapping of the relevant sample sizes, applied to large available datasets to represent the “true, entire” population. This allows us to derive parameters to adjust prehistoric data from actual archaeological datasets, in contrast to parametric bootstrapping, since we resample directly from an alternative archaeological dataset, not from a dataset constructed from a parametric model.

When Gini coefficients are estimated from individual-level data (rather than household, our benchmark), we use a Bayesian empirical likelihood (BEL) approach to move from the individual Gini coefficient to the household Gini. We pair individuals into couples using a prior distribution based on historical information from a different, but plausibly similar site. Since a likelihood is not available in closed form, we replace it with its empirical estimate. Combining the prior information with this empirical likelihood we compute the posterior, which can be used to estimate the household Gini.

For the problem of the missing households without wealth, differing assets, and population scale adjustments, we specify a prior distribution informed by additional historical and archaeological data. As in the couples’ case, the likelihoods are not available in closed form, so we use their empirical approximation to construct the posterior distribution of the adjusted Gini.

We hypothesize that these sources of bias and imprecision define a relationship between the Gini coefficient that one can compute directly on the available wealth data (including those computed by other scholars), called here G_i, for any past society and date i, and our best estimate G_i^A, adjusted by the above biases and comparable across different time and space. We model this relationship by the following equation

$$G_i^A = G_i(\frac{1}{\lambda_i}\mu_i\delta_i)(1-z_i) + z_i + k_i, \quad \textbf{[1]}$$

where as is appropriate for each site and date (using methods to be explained in the sections that follow) , λ_i is the factor to adjust for the sample to population bias just mentioned, and μ_i, z_i, δ_i and k_i are the factors to adjust for, respectively, differences in the wealth ownership unit, inclusion of those without wealth missing from the dataset, differences in the wealth asset type, and differences in population scale. We use this relationship to adjust all the raw Gini coefficients in our dataset and transform them into (ideally) unbiased, representative, and comparable measures, adopting a hypothetical benchmark case with a common asset type (household wealth), unit of observation (household), and population size (1,000 households).

A probability distribution of Gini coefficients for a particular site and date (analogous to that shown in Fig. 1) is then estimated using SE-propagation methods, conditional on the errors arising from the sources identified in Eq. **1** being uncorrelated. For each site and date we use Eq. **1** to compute 10,000 values of the Gini coefficient, each based on the raw wealth data and a random draw from the probability distribution of each of the five adjustment parameters of Eq. **1**. These probability distributions are shown in Fig. 2. Table 1 lists the adjustments and additional data used implementing them.

We call our method BRIDGE, an acronym that summarizes our approach and provides a basis for scholars interested in building on our work to account for bias and uncertainty in prehistorical and historical data. Our method exploits different statistical approaches but it is inherently Bayesian in nature, as we use datasets to estimate the distributions of correction parameters and we then sample from these distributions to propagate uncertainty, although, due to the complexity of the likelihood, we replace it with its empirical counterpart. Our approach resembles BEL as we empirically approximate the likelihood from the data, although we do not use a formal BEL framework. We use bootstrapping in only one of the five adjustments (Table 1), our assessment of the uncertainty due to small sample size. Finally, we propagate the parameter’s distribution numerically through an error propagation model. The propagation generates the posterior distribution of the true Gini coefficient and the associated 95% credible interval.

A key feature of our method is that when we estimate the parameters for the adjustments, we draw from archaeological datasets that are different from the ones whose inequality measure we adjust. Our approach is, therefore, forward-looking: given an observed Gini and independently estimated adjustment distributions, we propagate uncertainty numerically to obtain the distribution of the true Gini. By using information from other distributions we differentiate ours from other methods that derive the parameters by iterating the observations in the samples to be adjusted.

By using error propagation techniques, our work also broadly relates to recent developments in environmental research, where similar models are employed to estimate how error and uncertainties spread through complex and interrelated environmental, climatic, and hydrologic models (23–25). Our context and data structures are different than these cases, but our work similarly suggests that implementing an error propagation model captures a larger degree of uncertainty than the one that would be

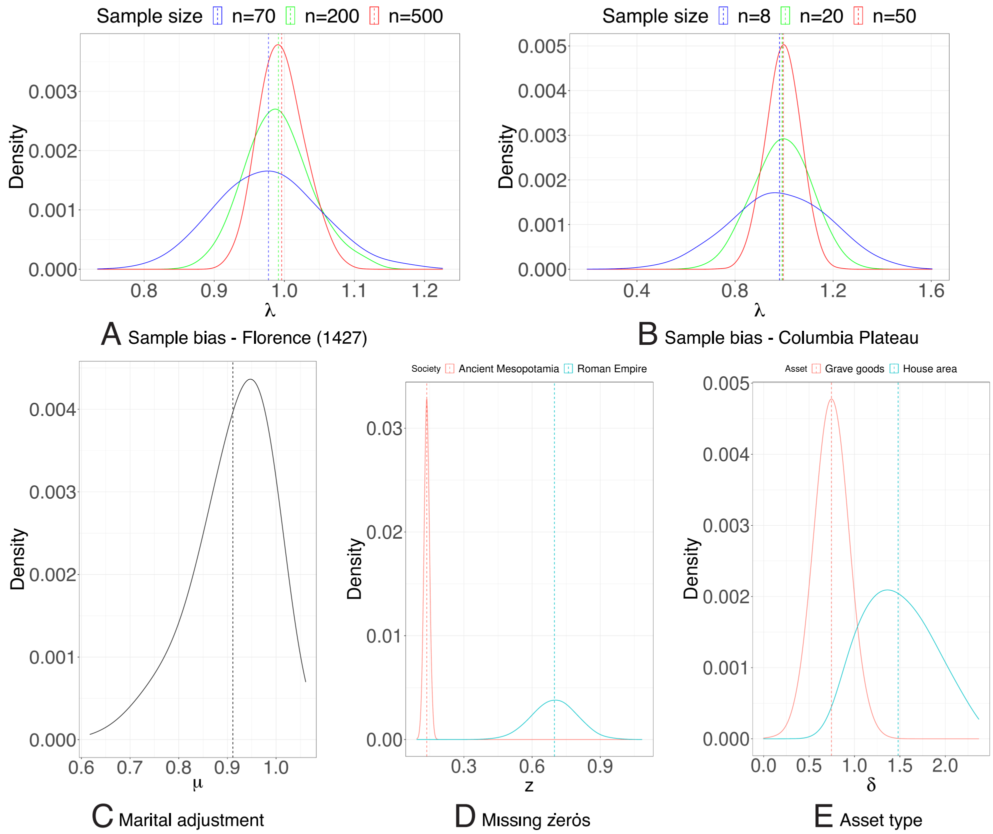


Fig. 2. Uncertainty of the raw data and adjustment factors. (*A* and *B*) Frequency distributions of the sample bias adjustment factors from the 1427 Florence and Columbia Plateau datasets. Dashed lines: mean value at each sample size. (*C*) Probability distribution of the marital adjustment factor and its mean, both computed merging the four archaeological cases. (*D*) Probability distributions for the fraction of missing zeros used for Ancient Mesopotamia (red) and Roman Empire (blue). Dashed lines: mean values of each distribution. (*E*) Probability distributions for the asset types factor to convert grave goods inequalities and house area into household wealth inequalities.

obtained by looking only at the errors from single stages of the analysis.

Bias and Imprecision due to Small Sample Size (λ)

We start by exploring the degree of uncertainty and bias in estimates of wealth inequality that would arise from random sampling in what are often very small-n datasets. We are motivated by a challenge faced by archaeologists: the fact that the materials they recover from a site provide evidence on a typically relatively small sample of a larger population that has not been studied or perhaps even discovered. Excepting the past century or so, which in many countries saw the virtually universal taxation of households, this problem is likewise common in historical sources. With few exceptions (one is the survey of the wealth of all households by the tax authorities in 15th century Florence that we use below) records of households' wealth are very restricted (either deliberately or unavoidably) to relatively small samples of the population of interest.

Table 1. Summary of adjustments and sources

Adjustment	Parameter in Eq. 1	Section	Method	Data used	Adjustment applied to
Sample Size	λ	3	Bootstrapping	Hohokam 11th c. CE; Florence 1427 CE; Columbia Plateau 10th–18th c. CE; Krummhörn 17th–18th c. CE	All estimates with sample size smaller than 10,000 (n = 380)
Individual to household ownership unit	μ	4	BEL	Columbia Plateau 10th–18th c. CE; Krummhörn 17th–18th c. CE	All estimates computed on individual-level (rather than household) data (n = 48)
Asset type	δ	7	BEL	Western Balkans (Gomolava) 5th millennium BCE; South Mesopotamia (Kafajah and UR) 2nd millennium BCE	House size and burial goods data (n = 204)
Including missing zeros	z	5	BEL	See *SI Appendix* and text	All datasets with missing zeros (n = 127)
Scale	k	6	BEL	Hohokam 11th CE; Columbia Plateau 10th–18th CE; Finland 1571 CE;	All estimates computed on sites with less than 250 households (n = 240)

*BEL: Bayesian Empirical Likelihood estimation.

To address the uncertainties and biases arising due to small samples, we make use of large datasets like the wealth data from Florence, and their analogues in the archaeological record. We suppose that the complete set of observations, Ω (of size M), is the whole population, and the Gini coefficient computed on Ω is the true Gini coefficient, $G(\Omega)$. We then repeatedly restrict the dataset to a set ω_m (of size m, where $m < M$) of randomly selected "found" individuals, and then estimate the Gini coefficient that would result ($G(\omega_m)$), using a large number of subsets of the data ω of size m. Where m_{min} is the size of the smallest of our datasets, for every $m_{min} \leq m < M$, we repeat a random sampling with replacement 10,000 times and generate a corresponding distribution of adjustment factors $\lambda(\omega_m) = G(\omega_m)/G(\Omega)$. To generate the probability distribution of each case in our dataset based on a sample of size m, we adjust the raw data estimate using a $\lambda(\omega_m)$ drawn from the distribution of λ's calculated in this way.

To take account of the difference in scale and economic structure between the societies on which we have archaeological as opposed to historical evidence, we do this exercise on two datasets, one historical and the other archaeological. The first is the extraordinary dataset on wealth distribution across 7,997 households in 1427 Florence (26). The distributions of adjustment factors $\lambda(\omega_m)$ for three different sample sizes shown in Fig. 2*A* indicate that a sample even as small as 70 (of the 7,997 in the hypothesized "entire" population of Florentine wealth holders) returns a remarkably small bias with a mean 98% of the "true" value and with the CI* intervals [83%, 120%]. When the sample size increases by a modest amount (e.g., to just $m = 200$) the bias remains (estimated mean is 99% of the true), but the estimate is substantially less uncertain (CI* [90%, 109%]). As a robustness check, we confirm in *SI Appendix* (Section A) that the method returns similar results when implemented on another historical dataset for the 17th–18th century German agricultural population at Krummhörn (27).

Our second, archaeological, dataset records burial wealth among 10th–18th century CE populations of Columbia Plateau (U.S) sedentary fishers (28). Fig. 2*B* shows that even a very small sample (e.g., $m = 8$) yields on average a modest level of bias but a substantial degree of uncertainty (mean 97% of the true, CI* [61%, 133%]). For somewhat larger samples (i.e., $m = 20$), the bias is smaller and the estimate less uncertain (mean = 99% of the true, CI* [77%,122%]). As a robustness check, we confirm in *SI Appendix* (Section A) that we arrive at similar results using a smaller archaeological dataset on the 11th century CE Hohokam culture of horticulturalists at la Ciudad, in modern day Arizona, U.S. (29).

The similarity of the ratios when estimated from a different population does not exclude the possibility that they might depend on the parameters of the distribution of the population. The bias or degree of uncertainty might be greater in highly unequal distributions, for example. In *SI Appendix* (Section A), we show that simulated log-normal distributions with different SD at the same population size exhibit similar distribution of the estimated λ parameters.

We use the distribution of $\lambda(\omega_m)$ from the Columbia Plateau dataset to sample-adjust each of the estimates based the archaeological data, while the historical data are sample-adjusted using the distributions estimated from the 1427 Florentine tax register. Thus our adjustment procedure, described below, allows us, and future scholars, to implement different parameters depending on whether the Gini to be adjusted is for a prehistoric or a historical society.

Estimating Household Wealth with Individual Level Data (μ)

When we have wealth data on individual men and women but do not know which individuals were paired in households, as it is often the case with burial goods, we create fictive couples based on alternative assumptions concerning marital assortment and calculate the wealth inequality measure on those units. The data for this exercise are from four pre-European contact North American burial sites with a substantial number of sex-identified individuals.

We assume that the members of couples are not buried in different sites, and that there is some degree of positive assortment on wealth in the pairing process. To create couples we rank all members of both sexes at a particular burial site by grave goods wealth level. We then generate wealth-assorted distributions with probability 0.5 matching a male with the female of identical rank (so that the n-th ranked male is matched with the n-th ranked female) and then matching randomly the remaining individuals who are not rank-matched by this process. We repeat the exercise 10,000 times at each of the four burial sites, each time computing the Gini coefficient for the unique set of fictive couples that we have created, and its ratio to the individual-based Gini coefficient.

These ratios, denoted μ, are the factors that we will use to convert individual grave goods inequality measures to the household benchmark. As can be seen from Fig. 2*C*, the couple-adjusted level of wealth inequality is substantially less than individual level inequality (mean $\mu = 0.910$), is quite precisely estimated (CI* [0.751, 1.014]), and the distribution is markedly left skewed. (The statistical details of the μ estimated from each single site are in *SI Appendix* (Section B). The degree of marital wealth assortment measured by average simple correlation in the wealth levels of the members of the fictive couples across the 10,000 randomizations for each of the four burial sites is 0.510, which is, not surprisingly, very close to the 50% of the "couples" for which we implemented perfect assortation.

Marital assortment by wealth is likely to be greater in economies in which material wealth is a more important factor of production (capital intensive farming for example). But lacking sufficient evidence of this, we apply the above estimated adjustment factor to all cases where a conversion from individual male and female wealth data to household wealth data is required (for the most part sedentary hunter-gatherers and labor limited farming prior to the introduction of ox-drawn plows).

As a robustness check we use the same method to calculate μ directly in a single dataset of semisedentary hunter-gatherers for which we know the separate wealth of the (living) male and female. We find a value of $\mu = 0.76$, which is consistent with our estimate that couple-adjusted wealth inequality is substantially less than individual-level inequality and also with our supposition that labor limited economies (like this hunter-gatherer population) may be less maritally assorted, constituting a bias in our adjustments that we are unable to address.

Comparability Across Asset Types and Other Wealth Indicators (δ)

The wealth of a household may be considered as a vector of assets of particular types: land, tools, housing, stored goods, domesticated animals, and so on. What we observe in ancient (and much historical) data is information on just one of these elements in the wealth vector, in many cases land, or some other measure thought to be a good proxy for aggregate wealth, such as burial goods.

To provide a single scalar measure of the household's total wealth, modern measures of net worth aggregate the distinct assets using actual or imputed market prices. But in many of the economies under study we either do not have information on market prices (for example, prior to the development of written records) or there are no markets for the exchange of the type of assets in question (as is often the case for land) and hence no prices. Even where some price evidence is available, there is little reason to believe that the then-existing forms of exchange sufficiently resembled competitive markets to take relative prices as plausible measures of the value of an asset.

In some cases—aggregating the heterogeneous items making up a grave ensemble, for example—archaeologists have used estimates of the labor time required to produce a good as a substitute for a price. Recent ethnographic data also allow estimates of shadow prices as the basis of aggregating heterogeneous assets making up a household's wealth. For example, shadow prices of cattle and land among the Kipsigis people in Kenya have been derived from the econometrically estimated relative contribution of these two forms of wealth to the reproductive success of their owner (taking large families as a value to which wealth contributes), allowing the aggregation of these two forms of wealth (30, 31). But these methods are not available for most of our sites and dates.

Typically we have a single measure, house size or grave goods, for example, but not both. There is reason to think that some forms of wealth will be systematically more unequally distributed than others, a source of bias that we illustrate by the case of grave goods. If burying valued goods with the deceased is a form of social signaling subject to constant or (more plausibly) increasing marginal returns and there are diminishing marginal benefits of other uses of wealth (for example in the form of housing) then it is likely that grave goods will be more unequally distributed than the total wealth of the living population. Because it is the latter that is of interest (the grave goods are just an indicator of total wealth) we would overestimate the degree of total wealth inequality by using unadjusted measures of grave goods.(We model the relationship between the two inequalities in a hypothetical population in *SI Appendix* (Section C).)

This same disparity between inequality in wealth and inequality in proxies for wealth appears to be present in other kinds of observations as well. The likely sharply diminishing marginal utility of additional house size beyond a certain point probably accounts for the evidence that living area of homes being much less unequally distributed than storage areas our benchmark form of wealth (32, 33, table 8.2).

To develop comparable measures for the different asset types we exploit cases in which we have measures of multiple forms of wealth in a single population (the details of the cases used are in *SI Appendix* (Section D)). This allows us to estimate δ_i, which is the ratio of the Gini coefficient for our common benchmark asset type household wealth (measured in our archaeological data by storage area) to the Gini coefficient for asset type i. Our adjusted Gini coefficient for asset i is then the observed measure times δ_i. For example, in Kafajah, South Mesopotamia, in the Early Dynastic Period (2900–2300 BCE) the Gini for grave goods is 0.779, the Gini for storage area (the benchmark wealth type) is 0.540 and hence $\delta = 0.693$, quantities very similar to the illustration based on a lognormal wealth distribution and the model of grave goods as signaling in *SI Appendix*, Fig. S4.

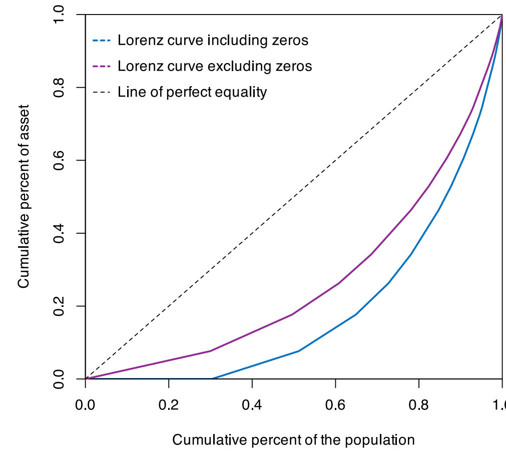


Fig. 3. The "missing zeros" problem: complete dataset and a hypothetical truncated dataset with missing zeros. Lorenz curves for all the n = 498 observations in the Columbia Plateau dataset when the zero wealth owners (30.32% of the dataset) are included (blue) and excluded (purple) from the dataset. Gini coefficients: true dataset (0.622), truncated dataset (0.457).

We use the mean and SD of the implied δ's estimated from the datasets in which we have Gini coefficients for multiple types of assets to generate a Gaussian prior distribution of probable values of asset type adjustments. As illustrations, Fig. 2*E* shows the distribution of δ's estimated in this way for converting grave goods and total house area measures to household wealth based measures: grave goods estimates are adjusted downward on average by 26% of the observed value ($\bar{\delta}_{grave} = 0.74$, CI* 40%; 111%), while house size estimates are adjusted upward by an average of 52% ($\bar{\delta}_{house} = 1.52$, CI* 82%; 214%). These are major adjustments with substantial CI, contributing significantly to our adjustments and their associated uncertainties.

Including Those Without Wealth Absent from Datasets (*z*)

A common problem in the historical and prehistoric measurement of inequality is that a source presents a measure of inequality of income, wealth, or some other dimension in which those without the attribute in question, those with zero income or the landless for example, do not appear. Were the raw data available and one knew the number of missing observations one could just recalculate, having inserted the missing zeros, yielding the desired measure. But in many cases, the researcher does not have access to the raw data including the missing households without wealth.

The problem is illustrated by the two Lorenz curves in Fig. 3, one calculated on the grave wealth of entire sample and the other a truncated Lorenz curve based on a sample from which we have hypothetically eliminated those without this form of wealth. Because the missing fraction without wealth is considerable, the underestimate of the Gini coefficient using the truncated dataset is substantial.

In *SI Appendix* (Section D) we show that if we have data on or can reliably estimate the number of excluded households without wealth, then an approximation of the true Gini coefficient (G^T) is possible, even if we do not have access to the raw data on which the truncated Gini coefficient was computed. It is remarkably simple: a weighted average of 1 and the Gini coefficient measured on the truncated population (G°) with the weights respectively being the fraction of the population without wealth (z) and its complement. So we have (using Eq. **1**):

$$G_i^o = G_i(\frac{1}{\lambda_i}\mu_i\delta_i)$$
$$G_i^T = z_i + G_i^{\circ}(1 - z_i). \quad [2]$$

Eq. **2** is an exact expression of the Gini coefficient for a hypothetical economy with just three wealth classes—the excluded without wealth, middling wealth, and substantial wealth—and an infinite population. It provides a nearly exact estimate of the Gini coefficient for the continuous wealth distributions in our dataset. We used 33 complete datasets from which we hypothetically removed the zeros, and then used Eq. **2** to estimate the true Gini coefficient. The mean absolute error between the true Gini and that estimated using Eq. **2** on the hypothetically truncated datasets, expressed as a fraction of the mean true Gini, is 0.00004 and the correlation coefficient between the two sets of values is 0.99 (the details are shown in *SI Appendix* (Section D)).

A further source of error is that in most of the historical cases, a substantial degree of uncertainty exists regarding the fraction of missing zeros in the population. Indeed, in almost all the cases for which our missing-zeros adjustment is implemented, the sources provide at best a range for the likely values of z. The best we can do is an educated guess based on existing sources, using these ranges to generate a Gaussian distribution of missing zeros, the mean of which is the range midpoint and the SD of which is selected so that some given share of the distribution's mass is outside the limits of the range (we arbitrarily used 10%).

In Fig. 2*D* we show two such distributions, one, for Bronze Age Mesopotamia (4th–2nd millennia BCE) in which our limited evidence suggests that the small numbers of missing zeros may be subject to modest uncertainty, and the other for the Roman Empire (3rd century BCE–4th century CE) for which a large fraction of missing zeros appears to be subject to substantial uncertainty. We use these parameters to generate a distribution of 10,000 z's from which draws are used along with Eq. **2** to estimate the true Gini coefficients (the historical sources and information used to bound the number of missing zeros in different contexts are described in *SI Appendix* (Section F)).

We adopt a similar approach when we have good reason to believe that a set of data (dwellings excavated by an archaeological team, for example) is unrepresentative of the underlying population (the entire settlement). If it is likely that the less wealthy households are those that do not appear in the data and if we can estimate the size of the entire population, we can reconstruct a hypothetical complete population using available data on the less wealthy portion of the population. See *SI Appendix* (Section F).

Scale Effects: Accounting for Differences in Population Size (k)

In many cases we have data on wealth inequality in just one of some lower level units ("villages") constituting a higher level unit (the "district"), the population of which is our hypothetical common benchmark size, so we need an estimate of inequality for the district. A scale effect arises because differences in mean wealth among the villages contribute to inequality at the district level. As a result estimates based on single settlements (the vast majority of our prehistoric data) will tend to understate the level of inequality at the district level. 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 sampled population in the late prehistoric phase is 0.642 while the average of the Gini coefficients for the six burial sites in this phase is 0.607.

To make our estimates comparable with respect to population, we seek to isolate pure scale effects due to differences in mean wealth levels among lower level units from unobserved other influences on inequality that are correlated with scale, such as the nature of the system of government (e.g., presence or absence of a state) or economy (producing food rather than hunting-gathering it). To do this we use independent information from similar populations on which we have complete information to estimate the difference between the Gini coefficients for the lower level population entities and the particular larger entity that they make up, as illustrated just above by the data from the Columbia Plateau. These differences, along with the difference in population of the lower units and the upper unit gives us an estimate of the effect of population size on measured inequality, which we can then use in cases where data is available on just a single village.[‡]

The advantage of this method is that it provides an estimate of the scale effect based on groups that, aside from differences in scale, are otherwise similar (for example, in political and economic structure) because the larger unit is composed of the smaller units. The scale effect estimated for a village or other lower level entity i, termed k_i, is given by

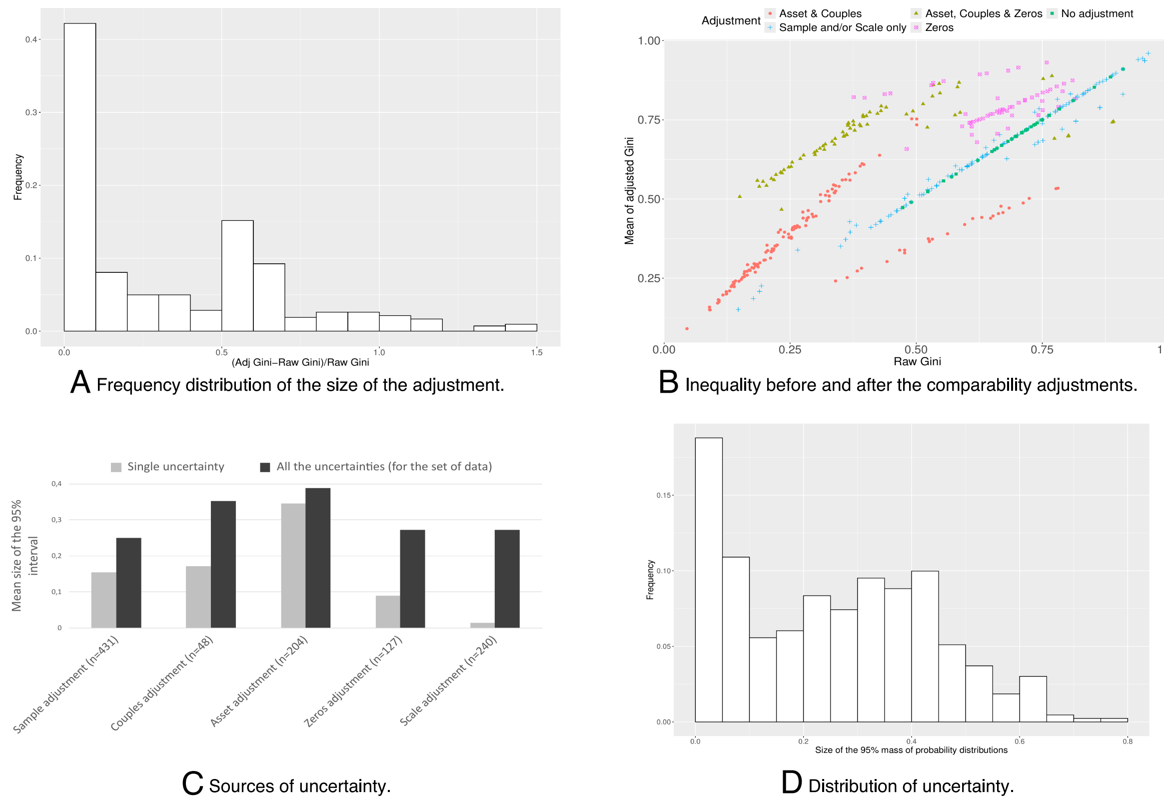
$$k_i = \frac{G_{j-}G_{ij}}{m_j - m_{ij}}, \quad [3]$$

where G_j and G_{ij} are the Gini coefficients respectively for the higher level entity and the lower level entity that is a constituent part of the higher level entity, and m_j and m_{ij} are the respective population sizes. We expect that the size of the scale effect will depend on the population size of the lower level entity.

To scale adjust our data we proceed in three steps. We first use datasets with nested structures of lower and higher level entities to estimate a large number of k_i's. Then we use these to estimate a statistical relationship between the magnitude of the individual scale effect and the population of the lower level entity giving us $\hat{k}(m)$, the estimated scale effect for a unit of size m. Finally, where we lack data on the higher level entities (the population level of interest) we use this estimated relationship to scale adjust the estimated Gini coefficients of the lower level entities. We call this the "nested method," and it corresponds exactly to our other "not found" thought experiments in bringing to bear independent data to assess the possible biases and imprecision in the estimates made on the basis of less than ideal datasets.

To implement our first step, there are three datasets that are nested in the above sense, allowing us to estimate the k_i's: the Columbia Plateau (28) and Hohokam archaeological datasets (29), and historical datasets for Finland in 1571 (35). Using the Finnish dataset as an illustration, in *SI Appendix*, Fig. S7 *A* and *B* we see that for each parish of one of the districts the scale effect is the slope of the line connecting an observation at the parish level with the observation at the level of the district of which the parishes were part. We estimate a total of 127 scale effects in this way. We used the Columbia Plateau and Hohokam

[‡]A natural way to make this adjustment might appear to be decomposing the Gini for an entire district on which we have data into a within village and between village component and use the ratio of the between to the total to make adjustments on data from sites with less complete data. But as has been known since (34), there is no proper decomposition of the Gini coefficient into within and between subunits Gini coefficients.



A Frequency distribution of the size of the adjustment.

B Inequality before and after the comparability adjustments.

C Sources of uncertainty.

D Distribution of uncertainty.

Fig. 4. Adjustment and uncertainty. (*A*) Distribution of the absolute value of the difference between adjusted and raw Gini as a fraction of the latter. (*B*) Mean of the probability distribution of likely Gini coefficients for each site and date (y-axis) and the raw Ginis (x-axis). Different shapes and colors indicate different adjustments implemented on the Ginis (depending on the nature of the data). All the Ginis except those in green have been adjusted by sample bias and/or scale effect. Those in red, pink, and yellow have also been adjusted by, respectively, asset type, missing zeros, and both. The cases shown in blue have been adjusted only by sample and or scale, which unsurprisingly in light of what we have found, result in only minor adjustments. (*C*) Numbers of cases with uncertainty in the type indicated in parentheses. Sample bias is implemented on 380 Ginis in our dataset, and the other four sources of imprecision are present (or not) depending on the characteristics of the data used to compute each raw Gini. Gray bars: extent of the uncertainty for only one source of uncertainty, giving the least span containing 95 % of the mass of the probability distribution of the adjusted Ginis. Black bars: size of the CI* (similarly defined) for the full adjustment. (*D*) Distribution of CI* for each estimate in the dataset. CI* distribution; mean = 0.367 (archaeological), 0.101 (historical).

archaeological datasets and Finnish data to estimate scale effects for respectively smaller and larger populations.

For the second step the estimated scale effects at each population level are then used to derive a function reflecting what we call the "estimated pure scale effect" showing how the Gini varies over the range of population sizes (*SI Appendix*, Fig. S7*C*). For purposes of our adjustment what matters is its slope, not the value of the Gini for any population level. So, we arbitrarily chose as a starting value of our function the Gini predicted by the nonnested function, estimated on our raw dataset, at 1,000 households. Then, from this arbitrary benchmark and the estimated slopes in *SI Appendix*, Fig. S7*C* we construct the estimated pure scale effect curve in *SI Appendix*, Fig. S7*D*.

SI Appendix, Fig. S7*C* shows that a very imprecisely estimated and substantial scale effect obtains for very small populations while a quite precisely estimated effectively zero scale effect holds for populations with more than 100 households. We incorporate this uncertainty in our adjustment method and use the SD of the estimated pure scale effect to construct a distribution of scale adjustment factors to be used in our adjustment process.

To illustrate the third and final step in our estimation, we show in *SI Appendix*, Fig. S7*D* one example of how the estimated pure scale effect function is used to size adjust the Gini coefficient originally estimated at a population level smaller than the benchmark value of 1,000 households used in our analysis (the method works in the same way for Gini coefficients estimated at population levels larger than 1,000 households.) The figure shows the adjustment of the Gini estimated for the late prehistoric period on the Columbia Plateau (after the sample bias and the demographic and asset adjustments described in the previous sections.)

A robustness check for our methods is provided by a recent paper coauthored by one the current authors that estimates the scale effect when using several different archaeological datasets from different world regions and periods. The results suggest that the scale effect behaves similarly across widely different areas and periods. They confirm that the scale effect captures the effect of size per se on inequality controlling for how other factors, e.g., political characteristics of larger units, might influence how size affected inequality. Scale effects estimated from different datasets also confirm that the effect depends mostly on the size of the small-level entity: the smaller the size, the larger the scale effect on inequality (36).

Results: A Taxonomy of Uncertainty

This completes our explanation of the estimation of the distributions of the five adjustment factors in Table 1 and Eq. **1**. To account for the uncertainty implied by the estimation of the parameters, our adjustment algorithm generates for each Gini coefficient 10,000 realizations of the computation described in Eq. **1** based on the raw data and a draw from each of the five

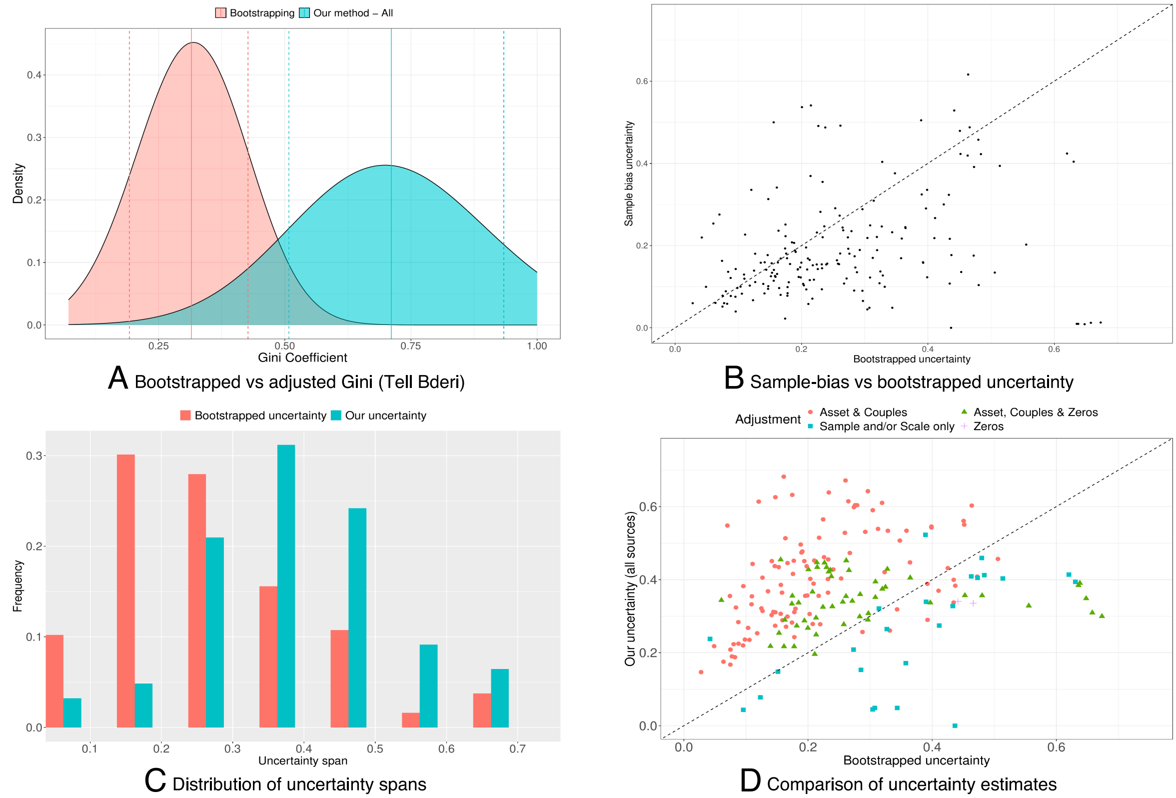


Fig. 5. Assessing the difference of methods to estimate uncertainty. (*A*) Distribution of the bootstrapped Gini coefficients (pink) with the one of the fully adjusted Gini according to our method (light blue) in Tell Bderi, Mesopotamia, 3rd millennium BCE. Means of CI*: bootstrapped distribution 0.314 [0.192;0.427], our distribution 0.711 [0.508;0.934]. (*B*) Relationship of the spans of the CI* estimated through the sample bias adjustment and bootstrapping. (*C*) Distribution of the spans estimated through the full adjustment and bootstrapping. (*D*) Relationship between the span of the CI* computed through our entire method of adjustment (y-axis) and the span of the CI* computed through bootstrapping methods (x-axis). The shapes and colors reflect the different combination of adjustments as shown in (*B*) of Fig. 4.

distributions of the five parameters. The result is a probability distribution of what we estimate to be the likely values of each true (meaning adjusted and comparable) Gini coefficient.

Looking across the entire dataset two things are clear. The first is that achieving comparability among our estimates requires substantial adjustments. Gini coefficients estimated from data excluding those without the asset in question or on differing types of assets without a comparability adjustment, for example, will not be very informative about underlying wealth inequalities. The second is that for most of the estimates, the degree of uncertainty in the estimates is substantially greater than what one would infer from conventional bootstrapped estimates.

The Propagation of Uncertainty: An Illustration. We first explain our procedure by illustrating the adjustment of the Gini coefficient for Akkadian Mesopotamia (2350–2150 BCE) the probability distribution of Gini coefficients for which was shown in Fig. 1.

Our algorithm works in this way. To account for the bias and degree of uncertainty of our estimate due to sampling, the raw Gini coefficient for the Akkadian Mesopotamia (0.802), computed on a sample of 206 graves, is first multiplied by 10,000 instances of the $\lambda(\omega_m)$ parameter estimated from the Columbia Plateau dataset at $m = 206$. This first step transforms the reported Gini coefficient into a distribution of 10,000 Gini coefficients reflecting our estimate of both the imprecision and bias of the sample. The mean of these is 0.803, CI* [0.760; 0.848], there being virtually no adjustment due to the substantial sample size of this estimate.

Then, as this estimate was computed on individuals' grave goods (22) rather than couples' (or households'), we implement the individual-to-couple wealth adjustment. We do so by multiplying 10,000 random draws (with replacement) from the sample-size-adjusted Gini coefficients each by a level of the parameter μ randomly drawn (with replacement) from the distribution shown in Fig. 2*C*. We obtain the new distribution of the 10,000 now sample-and-couple-adjusted Gini coefficients (with mean 0.731, CI* [0.603;0.814]).

We implement this procedure for all the subsequent adjustments encompassed by Eq. **1** and we finally obtain a distribution of 10,000 Gini coefficients that we consider our best estimate of the probability distribution of the unbiased and comparable representation of the Gini coefficient for Akkadian Mesopotamia. Starting from a reported raw Gini of 0.802, the implementation of our model results in 10,000 fully adjusted Gini coefficients, with mean 0.700 and CI* [0.510;0.878]. In this case the adjustments to the raw estimate result in a 13% reduction in the value of the Gini coefficient and CI* for the fully adjusted statistic that is more than four times as large as for the initial statistic computed above, that is, taking account of an estimate of sampling error alone.

Adjustments for Comparability Are Substantial. In Fig. 4*A* we show the frequency distribution of the size of the adjustment (normalized by the magnitude of the raw observation.) 44% of the adjustments are less than 15% of the magnitude of the raw data, but there is a small number of much larger comparability adjustments. The mean absolute value of these adjustments is 0.361 of the mean of the unadjusted Gini coefficients.

Fig. 4*B* shows the relationship between the raw estimates and the mean of the probability distribution of the fully adjusted coefficients. Most of the Gini coefficients that are adjusted only by sample and scale adjustment (shown in blue) differ at most modestly from the raw data. The Gini coefficients with more substantial comparability adjustments are those that are also adjusted by the individuals to couples adjustment, the missing zeros adjustment, and the asset adjustment for grave goods and for house size. The Pearson correlation coefficient between raw and adjusted Gini coefficients is 0.790.

Sources of Uncertainty. We can assess the relative importance of the various sources of uncertainty—sample bias, adjustment from individuals to households, asset type adjustment, inclusion of missing zeros, and scale differences—across our entire dataset by calculating CI* for the distribution Gini coefficients for each site and date when just a single source of uncertainty is included (effectively "turning off" the remaining sources.) Fig. 4*C* presents the results of this exercise.

Each pair of bars concerns a particular source of uncertainty (sampling error, or uncertainty in the adjustments for comparability). Because these sources of uncertainty pertain to differing subsets of the data, the number of cases and the aggregate uncertainty (from all sources for the particular subset of the data) differ across the sets of bars. We can see that for the 204 cases in which an asset type adjustment was warranted, uncertainty concerning the adjustment constitutes most of the uncertainty in the dispersion of the distribution of full adjusted Gini coefficients. By contrast, for the 240 cases in which a population size adjustment was required, that adjustment made a trivial contribution to the overall uncertainty of the estimates.

The Extent of Uncertainty Is Substantial and Varies by Source. In Fig. 4*D* we provide evidence on the degree of uncertainty, showing the distribution of the CI*, the least spans containing 95% of the mass of the relevant distribution of the adjusted Gini coefficients in the dataset. These intervals are less than a third of the mean of the probability estimate in 40% of the cases, but intervals of half the magnitude of the mean or greater are also observed in a substantial fraction of cases (45%). The mean degree of uncertainty measured in this way (CI* divided by the mean) is 0.532.

Comparing Our Measure of Uncertainty with Alternative Methods. The historical and archaeological studies of inequality that take explicit account of the degree of uncertainty often use bootstrapping, that is, repeated sampling with replacement from the raw data to generate a large number of alternative statistics for the quantity of interest, the variance (or other measure of dispersion) of which is the measure of uncertainty (33, 37–39).

In Fig. 5*A* we illustrate our results by the case of a 3rd millennium BCE Mesopotamian settlement, comparing our fully adjusted distribution of Gini coefficients with a bootstrapped distribution. The difference in the mean of the distributions is due to the fact that we have adjusted the data, with (as would be expected from the previous section) a substantial increase coming from the fact that the raw data are on house size while our benchmark asset type is storage area, which typically is substantially more unequally distributed. The much wider CI* (75% greater) in our fully adjusted distribution is due (as expected) to uncertainties that we have estimated beyond sampling error in the generation of the final data.

A more general comparison of the CI* obtained with our method with those from bootstrapping method for the 186 cases for which we have the raw data is shown in the other three panels of Fig. 5. We first compare the uncertainty measured by bootstrapping, with the one obtained when implementing our sample bias adjustment. The reason for such comparison is that the two methods address uncertainty rising from sample composition. The average CI*'s obtained from sample bias adjustment is 0.191 [0.039;0.424], while the one from bootstrapping is 0.258 [0.074;0.473], with a p^* for the null hypothesis that the bootstrapping uncertainty measures are not larger than our sample bias uncertainty (<0.001). Panel (*B*) plots the two CI*'s for each of the 186 cases. The spans of the two CI*s–based on our adjustment and the bootstrap–have a correlation coefficient equal to 0.269 (*p*-value 0.0001), and they show that the bootstrapped estimates of uncertainty are largely uninformative about the degree of uncertainty as we have measured it. It is also clear that bootstrapping. techniques overestimate the uncertainty regarding the sample bias.

In contrast, Panel (*C*) shows that the measures of uncertainty from our fully adjusted distribution are generally larger than the ones obtained from bootstrapping: the mean of our uncertainty spans is 1.42 times the mean of the bootstrapped measures of uncertainty (p^* for the null hypothesis that our uncertainty measures are not larger than the bootstrapped measures is 0.0001). Panel (*D*) indicates that not only are the preponderance of our measures of uncertainty greater than the bootstrapped measures, but the bootstrapped measures are largely uninformative about the degree of uncertainty estimated as we have done (The simple correlation coefficient between our measure of uncertainty and the bootstrapped measure is equal to 0.189). It appears that measures of imprecision based on bootstrapping may give the appearance of accounting for our ignorance without adequately doing so.

But the underestimate of uncertainty is far from uniform. Importantly, the bootstrapped estimates substantially overstate the degree of uncertainty for cases in which the only adjustments concern small sample size and scale (the turquoise dots below the 45° line). This provides a further illustration of the value of exploiting additional data sources, in this case showing that relatively small random samples from a larger population provide remarkably precise and unbiased estimates of the "true" quantities based on the larger dataset.

Discussion: Beyond "Uncertainty Laundering"

The substantial degree of uncertainty in estimates of wealth inequality that we have identified here is unavoidable in studies of economic disparities not only in ancient times but even for more recent history. Conventional statistical tests of significance using single point estimates of the statistic of interest, we believe, constitute a form of what Andrew Gelman (40) called "uncertainty laundering" [or perhaps, "incredible certitude," the term used by Charles Manski (41).]

Our proposed alternative to these methods, BRIDGE, might at first sight seem to be a complex procedure, requiring a substantial amount of learning. But it presents at least four advantages. First, it is a fully replicable method, with an open access algorithm and datasets that can be implemented on any new available data point. Second, it can be easily modified, in each of its steps, and it can be augmented or reduced depending on the context. Third, it allows comparisons of the original raw data with the corrected ones and the associated CI. Finally, it provides CI that are not necessarily larger but are different from the ones that would be obtained by bootstrapping the raw Gini coefficient, depending on the types of adjustments required by the nature of the data.

Recognizing the difficulties intrinsic to aggregating differing forms of wealth without a set of prices as an indicator of their relative values, some have suggested treating wealth as irreducibly heterogeneous and using multiple measures (42), or adopting composite indexes based on principal components analysis or similar techniques (43).

The advance of research methods in archaeology—for example data acquisition assisted by LiDAR (Light Detection and Range) to create mini topographies of sites—will surely mitigate these uncertainties and possibly allow improved methods of uncertainty estimation. This is likely to be the case with respect to the sample bias problem. Recall that satellite-generated data from Ancient Mesopotamia returned a population of sites than differed dramatically from conventional excavation-based methods (22). If new technologies make possible a more precise understanding of the relationship between the excavated portion of a site and the whole underlying population, for example, a revised analysis could assess which parts of the population would be missing from a sample.

We are less confident that new archaeological methods will fundamentally alter the uncertainties with respect to the four other sources that we have addressed. The mapping from individual grave goods to household wealth will remain obscure. The comparison of wealth inequality among different assets will surely benefit from more evidence, but when data on only one asset are available, the uncertainty will remain regarding how the other assets were distributed in the population. Our adjustments will still be needed when those owning nothing are missing from datasets, with new techniques possibly providing better estimates of the numbers of missing zeros. Last, when comparing cases of different population scales, improved archaeological techniques will help in discovering the entire composition of larger administrative units, while our method will still be needed in cases where this remains impossible. In all of these cases, we believe that our method will remain important for the careful use of both prehistoric, historical, and even current data.

In response to the substantial sources of uncertainty that we have identified, we followed Bowley's advice and sought to "tabulate our ... ignorance" by means of error propagation methods and a robust accounting of the likely sources of bias and imprecision both in the raw data and in the independent sources for our representativeness and comparability adjustments. We have gone beyond shortcuts like bootstrapped SE by modeling the uncertainty generating process based on what is known or can be inferred about how the data came to be available and other independent information.

But our probability distributions of likely Gini coefficients for each site and date can hardly be said to be an adequate catalog of Morgenstern's "fact of error." Even modified by our simulation of more representative populations including missing households without wealth, for example, our "not found" exercises cannot fully account for the nonrandom nature of the archaeologists' or historians' processes of search, estimation, and publication. Moreover the nature of the household as a wealth sharing unit—the access of nonmembers to the benefits of a household's assets, for example—surely differs greatly across our sites and dates. (Evidence of this is provided in ref. 10.) Finally, our algorithm for adjusting the Gini coefficients—random draws from the probability distributions of the adjustment parameters—entails the assumption that errors in each of the adjustments are uncorrelated, which as an approximation may be generally plausible but could be significantly violated in some cases.

We can be reasonably confident, nonetheless, in the following. First, even very small random samples from a larger population yield a quite precise and substantially unbiased estimate of the Gini coefficient. Second, for plausible levels of marital assortment on wealth, household wealth is substantially less unequal, though the degree of overestimation entailed by using individual data is imprecisely estimated. Third, measures of wealth inequality based on different assets are not comparable without substantial adjustments (which with data currently available are subject to substantial imprecision).

Fourth, calculations of the Gini coefficient based on data that exclude those without wealth are substantially underestimated; but conditional on having reliable estimates of the number of those excluded, our proposed algorithm for simulating the complete distribution provides a remarkably accurate estimate. Fifth, except for very small populations, scale effects combined with incomplete estimates of inequality at a higher level unit (e.g., a district) based on estimates of a subset of the lower level units (villages) entail at most modest underestimates of the Gini coefficient at the higher level.

Combining independent information on the sources of uncertainty and bias in an error propagation framework is a general method that can be applied widely in archaeology, economics, history, and other fields. Like our estimates of wealth inequality, estimates of wages that are comparable across sites and dates, for example, are subject to uncertainties in the relative values of the currencies in which they are paid, the price level at differing dates, the hours of work, and the extent of in-kind payments, all of which propagate through the estimation process. We have suggested methods for implementing our overall strategy, hoping that these will be applied to other research domains and modified and improved by other users (we are currently developing a user-friendly tool for researchers interested in implementing our full method without additional coding).

Data, Materials, and Software Availability. M.F. and S.B., Data and code for "Accounting for uncertainty and bias in archaeological and historical evidence on wealth inequality" have been deposited in Inter-University Consortium for Political and Social Research, Ann Arbor, MI, 09 May 2026 (https://doi.org/10.3886/E248045V2) (44).

ACKNOWLEDGMENTS. We would like to thank Max Greenberg, Cosma Shalizi, Francesco Azzoni, Botond Szabo, Giacomo Zanella, Cody Ross, Simon DeDeo, Steven Durlauf, Amos Golan, and Justin Weltz for their comments on earlier drafts and to the Emerging Political Economies Project and Caroline Seigel (both at the Santa Fe Institute) for their support.

Author affiliations: [a]Department of Social and Political Sciences, Dondena, Centre for Research on Geography, Resources, Environment, Energy & Networks, and Bocconi Institute for Data Science and Analytics-Bocconi University, Milan 20136, Italy; [b]Centre for Economic Policy Research, London EC1R 5HL, United Kingdom; [c]Santa Fe Institute, Santa Fe, NM 87501; and [d]CORE Economics, University College London, London WC1HOAX, United Kingdom

1. B. Milanovic, P. Lindert, J. Williamson, Pre-industrial inequality. *Econ. J.* **121**, 255–272 (2010).
2. G. Alfani, Economic inequality in preindustrial times: Europe and beyond. *J. Econ. Lit.* **59**, 3–44 (2021).
3. J. L. Van Zanden, *The Long Road to the Industrial Revolution. The European Economy in Global Perspective* (Brill, 2009), pp. 1000–1800.
4. W. Scheidel, *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century* (Princeton University Press, Princeton, NJ, 2017).
5. T. A. Kohler *et al*., Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. *Nature* **551**, 619–22 (2017).
6. A. Bogaard, M. Fochesato, S. Bowles, The farming-inequality nexus: New methods and evidence from Western Eurasia. *Antiquity* **93**, 1129–1143 (2019).
7. M. Fochesato, A. Bogaard, S. Bowles, Comparing ancient inequalities: The challenges of comparability, bias and precision. *Antiquity* **93**, 853–869 (2019).
8. G. Dow, C. Reed, *Economic Prehistory: Six Transitions That Shaped the World* (Cambridge University Press, Cambridge, UK, 2023).
9. R. C. Allen, The neolithic revolution in the Middle East. *Econ. Hist. Rev.* **77**, 1154–1196 (2024).
10. S. Bowles, M. Fochesato, The origins of enduring economic inequality. *J. Econ. Lit.* **62**, 1475–1537 (2024).
11. T. A. Kohler *et al*., Economic inequality is fueled by population scale, land-limited production, and settlement hierarchies across the archaeological record. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2400691122 (2025).
12. O. Morgenstern, *On the Accuracy of Economic Observations* (Princeton University Press, Princeton, NJ, 1963).
13. A. L. Bowley, E. Cannan, F. Edgeworth, H. Lees Smith, W. Scott, The amount and distribution of income (other than wages) below the income tax exemption limit in the United Kingdom. *J. R. Stat. Soc.* **74**, 37–66 (1910).
14. W. C. Mitchell, W. I. King, F. R. Macaulay, O. W. Knauth, *Income in the United States: Its Amount and Distribution, 1909–1919, Volume 1: Summary* (Arno Press, New York, NY, 1921).
15. S. Kuznets, *National Income and Its Composition, 1919–1938* (National Bureau of Economic Research, New York, NY, 1941).
16. C. Feinstein, M. Thomas, A plea for errors. *Hist. Methods* **35**, 155–165 (2002).
17. A. L. Bowley, National income in America and the United Kingdom. *Economica* **9**, 27–36 (1942).
18. R. Stone, Two studies of income an expenditure in the United States. *Econ. J.* **53**, 60–75 (1943).
19. A. Bogaard *et al*., Labor, land, and the global dynamics of economic inequality. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2400694122 (2025).
20. S. Bowles, W. Carlin, Inequality as experienced difference: A reformulation of the Gini coefficient. *Econ. Lett.* **186**, 108789 (2020).
21. M. Fochesato, S. Bowles, "Technology, institutions, and wealth inequality: What do we learn from comparable estimates over 12 millennia?" (CEPR Discussion Papers No. 19064, 2024).
22. E. C. Stone, "The trajectory of social inequality in ancient Mesopotamia" in *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences, Amerind Seminar Series*, T. A. Kohler, M. E. Smith, Eds. (University of Arizona Press, Tucson, AZ, 2018), pp. 230–261.
23. G. Heuvelink, *Error Propagation in Environmental Modelling with GIS* (CRC Press, 1998).
24. E. Vogel *et al*., An evaluation framework for downscaling and bias correction in climate change impact studies. *J. Hydrol.* **622**, 129693 (2023).
25. H. Wu, X. Ye, B. Zhang, B. Chen, Assessment of uncertainty propagation from climate modeling to hydrologic forecasting under changing climatic conditions. *J. Environ. Inf.* **45**, 27–41 (2025).
26. D. Herlihy, C. Klapisch-Zuber, *Tuscans and Their Families. A Study of the Florentine Catasto of 1427* (Yale University Press, New Haven, CT, 1985).
27. K. Willfuhr, C. Stormer, Social strata differentials in reproductive behavior among agricultural families in the Krummhörn Region (East Frisia, 1720–1874). *Hist. Life Course Stud.* **2**, 58–85 (2015).
28. R. Schulting, *Mortuary Variability and Status Differentiation on the Columbia-Fraser Plateau* (Archaeology Press, Simon Fraser University, Burnaby, BC, Canada, 1995).
29. R. Mcguire, *Death, Society, and Ideology in a Hohokam Community* (Westview Press, Boulder, CO, 1992).
30. C. T. Ross *et al*., Greater wealth inequality, less polygyny: Rethinking the polygyny threshold model. *J. R. Soc. Interface* **15**, 20180035 (2018).
31. S.-Y. Oh, C. T. Ross, M. B. Mulder, S. Bowles, "The decline of polygyny: An interpretation (2017). https://www.santafe.edu/research/results/working-papers/decline-polygyny. Accessed 31 May 2026.
32. A. Bogaard, A. Styring, J. Whitlam, M. Fochesato, S. Bowles, "Farming, inequality and urbanization: A comparative analysis of Late Prehistoric Northern Mesopotamia and South-West Germany" in *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences*, T. A. Kohler, M. E. Smith, Eds. (University of Arizona Press, Tucson, AZ, 2018), pp. 201–229.
33. K. Ames, C. Grier, Inequality on the Pacific Northwest Coast of North America measured by house-floor area and storage capacity. *Antiquity* **94**, 1042–1059 (2020).
34. N. Bhattacharya, B. Mahalanobis, Regional disparities in household consumption in India. *J. Am. Stat. Assoc.* **62**, 143–161 (1967).
35. I. Nummela, "Regional inequality" in *Physical and Cultural Space in Pre-Industrial Europe. Methodological Approaches to Spatiality*, E. B. M. Lamberg, Ed. (Nordic Academic Press, 2011).
36. E. R. Crema, M. Fochesato, A. G. M. Ramón, J. Munson, S. G. Ortman, Toward multiscalar measures of inequality in archaeology. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2400700121 (2025).
37. P. M. Dixon, J. Weiner, T. Mitchell-Olds, R. Woodley, Bootstrapping the Gini coefficient of inequality. *Ecology* **68**, 1548–1551 (1987).
38. M. Biewen, Bootstrap inference for inequality, mobility and poverty measurement. *J. Econometrics* **108**, 317–342 (2002).
39. A. Squitieri, M. Altaweel, Empires and the acceleration of wealth inequality in the pre-Islamic near east: An archaeological approach. *Archaeol. Anthropol. Sci.* **14**, 190 (2022).
40. A. Gelman, The problem with p-values is not just with p-values. *Am. Stat.* **70**, 1–2 (2016).
41. C. F. Manski, The lure of incredible certitude. *Econ. Philos.* **36**, 216–245 (2020).
42. R. C. Oka *et al*., "Dreaming beyond the Gini: Methodological steps toward a composite archaeological inequality index" in *Ten Thousand Years of Inequality: The Archaeology of Wealth Differences*, T. A. Kohler, M. E. Smith, Eds. (University of Arizona Press, 2018), pp. 96–129.
43. A. M. Prentiss, T. A. Foor, A. Hampton, E. Ryan, M. Walsh, The evolution of material wealth-based inequality: The evidence from Housepit 54, Bridge River, British Columbia. *Am. Antiq.* **83**, 598–618 (2018).
44. M. Fochesato, S. Bowles, Data and code for "Accounting for uncertainty and bias in archaeological and historical evidence on wealth inequality." Inter-University Consortium for Political and Social Research. https://doi.org/10.3886/E248045V2. Deposited 9 May 2026.