

We Do Not Know the Population of Every Country in the World for the Past Two Thousand Years

TIMOTHY W. GUINNANE

Economists have reported results based on populations for every country in the world for the past two thousand years. The source, McEvedy and Jones' *Atlas of World Population History*, includes many estimates that are little more than guesses and that do not reflect research since 1978. McEvedy and Jones often infer population sizes from their view of a particular economy, making their estimates poor proxies for economic growth. Their rounding means their measurement error is not "classical." Some economists augment that error by disaggregating regions in unfounded ways. Econometric results that rest on McEvedy and Jones are unreliable.

"... we haven't just pulled the figures out of the sky. Well, not often."

—McEvedy and Jones (1978, p. 11)

In the past 20 years, economists have estimated empirical exercises that rely in part on a published work that reports the population of every country in the world starting in the year 1 CE or even earlier. The existence of such data surprises those familiar with research on population history; we have only a rough idea of the population of most parts of the globe before 1500. For many countries, the statistical lacuna extends closer to the present. Until the advent of modern censuses, which in most countries started during the nineteenth century, reckonings of the total population for even the best-studied cases remain subject to considerable error.

The Journal of Economic History, Vol. 83, No. 3 (September 2023). © The Author(s), 2023. Published by Cambridge University Press on behalf of the Economic History Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited. doi: 10.1017/S0022050723000293

Timothy W. Guinnane is Professor Emeritus, Department of Economics, Yale University, New Haven, CT 06520. E-mail: timothy.guinnane@yale.edu.

For comments and suggestions I thank Robert Allen, Cihan Artunç, Quamrul Ashraf, Gareth Austin, Emiliano Travieso Barrios, Howard Bodenhorn, Christophe Chamley, Tracy Dennison, Jeremy Edwards, Shari Eli, William English, James Fenske, Philip Hoffman, Maggie Jones, Thomas Mroz, Steven Nafziger, Tom Nicholas, Sheilagh Ogilvie, Paul Rhode, Mark Rozenzweig, Larry Samuelson, Richard Tilly, Christopher Udry, Warren Whatley, Qiyi Zhao, and seminar participants at Oxford and Yale. Benjamin English provided excellent research assistance.

These exercises typically rely on McEvedy and Jones's *Atlas of World Population History* (hereafter MJ). Published in 1978, this work reports a population total for the countries of the world at intervals of a century or half-century. MJ did not disguise the rough nature of their data, as the epigraph notes, and we should distinguish what they report from the way others used their work. Several economists point to a U.S. Census Bureau summary that appears to endorse MJ's estimates. The Bureau simply notes that MJ's estimates for *world* population are not too different from the other, earlier results.¹ As MJ state (pp. 353–4), however, that agreement is largely by construction.

The drawbacks of using such data are numerous. MJ's estimates, as they suggested themselves at the time of writing, lacked, in many cases, any firm foundation. Often, the estimates appear to reflect a judgment about the nature of the economy in question, rendering their use as economic proxies partially tautological. The MJ estimates are out-of-date for some countries; researchers have provided better figures in the past 40 years. Economists tend to dismiss measurement error issues by appealing to the implications of "classical" measurement error. MJ's clearly stated rounding rules mean the measurement error is not classical. Non-classical measurement error create several opportunities for bias in regression models. Economists have compounded these weaknesses with unwise disaggregation practices.

Many economics articles, including several highly cited contributions in the leading journals, rely on MJ for econometric exercises. This research has appeared in the leading general-interest economics journals, in development and growth-oriented journals, and in the main field journals for economic history. Several of these papers have been cited many times.² The present paper raises serious questions about the results of any econometric exercise that relies on MJ.

¹ Acemoğlu, Johnson, and Robinson (2002, p. 1242) and Ashraf and Galor (2011, note 14) each refer to a webpage that no longer exists. This page, <https://www.census.gov/data/tables/time-series/demo/international-programs/historical-est-worldpop.html>, lists only world population totals. Later, I discuss the other estimates.

² As of December 2021, Google Scholar lists more than 15 citations to MJ in the "Top 5" economic history journals, 8 in the *Journal of Development Economics*, 9 in the *Journal of Economic Growth*, and more than 25 in the four main English-language economic history journals. Some of the most highly cited empirical papers of the past 20 years rely on MJ. Examples include Acemoğlu, Johnson, and Robinson (2001), with more than 15,000 citations; Acemoğlu, Johnson, and Robinson (2002), with more than 5,000; and Acemoğlu and Johnson (2005), with more than 4,000. The MJ estimates are also a source for historical population in the database The History Database of the Global Environment (HYDE). See Goldewijk et al. (2017). While this paper focuses on economics, other historically-minded social scientists have also relied on MJ. See, for example, Baumard et al. (2015).

If the correct population data were available, we could re-estimate specific models that appear in published papers and assess the consequences of the measurement-error problems discussed here. This is obviously not possible because we lack the correct data. What I do instead is study the way MJ assemble and round their estimates. This permits us to draw on econometric literature to understand the difference between a model estimated using MJ and a model estimated using corrected population data. I then discuss more specifically the way some economists have used this population data. A brief replication exercise using Nunn and Qian (2011) shows that some published results are not robust to careful consideration of the problems in the MJ data.

THE SOURCE

MJ report a series of graphs of total population in a country (or region), with labels at centuries or half-centuries. Figure 1 reports the data for Germany in a format similar to the figures MJ use to present most of their estimates.³ For the twentieth century and, in some cases, the nineteenth, MJ reproduce official census counts as discussed by earlier scholars, sometimes adjusted for changes in national boundaries.⁴ Modern censuses did not start anywhere until the late eighteenth century and were not widespread until the nineteenth century.

One would think from reading the economics literature that MJ report precise numbers based on their analysis of earlier works. Graphs such as Figure 1, along with MJ's descriptions, suggest a different picture. "There are almost no data on which to base a population estimate for Germany until we reach the late Middle Ages" (McEvedy and Jones 1978, p. 70). "Estimates of Poland's population before the 14th century are based on nothing more than general ideas about likely [population – T.G.] densities" (p. 76). For the Maghreb, "There is really nothing on which to base any calculations before the 19th century" (p. 220). These comments are admirably frank, but MJ do, in fact, report population totals for Germany, Poland, and the Maghreb, and economists have used those observations to test hypotheses we view as important. MJ include a bibliography for each group of population estimates, but they typically do not explain how they used the references they list. For Burma, they

³ We were unable to secure permission from MJ's publisher to include their actual figures. Guinnane (2021, figure 1) is the relevant page from MJ. The replication package for this article is Guinnane (2023).

⁴ With exceptions discussed later, MJ rely on the definitions and boundaries of countries as of 1975. Imputing constant-territory population for countries whose borders change over time raises additional issues I do not address.

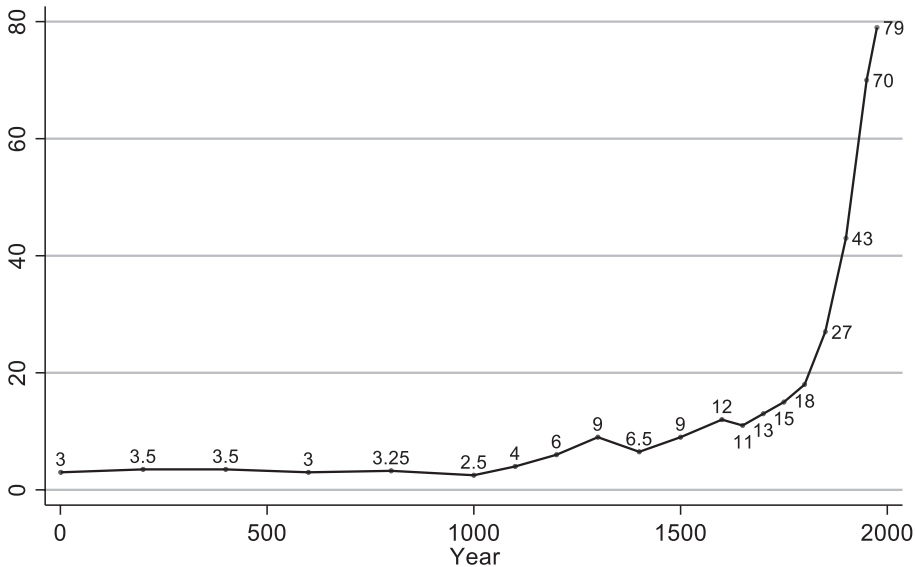


FIGURE 1
 THE POPULATION OF GERMANY 1 CE–1975
 ACCORDING TO McEVEDY AND JONES

Notes: Population figures are in millions, rounded by McEvedy and Jones as discussed in the text. Years are centuries until 1600, then include the years 1650, 1750, and 1850. The original figure is reproduced in Guinnane (2021, figure 1). I was unable to secure permission to reproduce that graph here.

Sources: McEvedy and Jones (1978, p. 69), from datafile provided by James Fenske.

note that the quantitative record consists of a single publication based on a count of houses in 1783 as well as colonial censuses that began in 1871. Yet MJ report population sizes for that country as far back as 400 BCE (pp. 190–92). This is not an isolated example. In discussing the western hemisphere, they refer to debates current at the time they wrote, but those debates suggested large ranges of estimates and pertain to the decades just prior to European contact. Yet MJ provide estimates for countries in this region going back many centuries. Most African entries have the same flavor; the only evidence MJ cites refers to the seventeenth century at the earliest, yet they report estimates for two full millennia.

Caldwell and Schindlmayr (2002, p. 200) emphasize the difficulty of useful population estimates for most of the world, and even for Europe before 1800. Population figures for the large areas of the globe that once fell under European colonial domination may be the hardest part of the problem. The essays on the Americas collected in Denevan (1992) document debates that continue. Carlos, Feir, and Redish (2022, p. 522), for example, note that in the early twentieth century, estimates of the

pre-contact population of North America north of Mexico City ranged from 1.2 to 18 million people. Later efforts narrowed that range to 1.2 to 6.1 million people. The demographic consequence of colonial contact is one measure of imperialism's impact on indigenous peoples, so these population measures carry considerable interpretative weight. Continuing differences of opinion do not reflect a lack of research.⁵

Systematic discussion of MJ has been limited, but specialists tend not to be impressed. As Austin (2008, p. 1102) puts it, "If you look up McEvedy and Jones expecting a treatise, detailing the original evidence and the reasoning behind the judgements by which it was converted into useable data, you will be disappointed." In discussing one particular study that relies heavily on MJ, Austin (2008, p. 1002) says that "there is simply no epistemological basis for Nunn's use of the word 'data' – literally, 'things that are given' or granted – to refer to the guesses that have been made about the population of future African countries in 1400."⁶

MJ's effort reflects a long interest in the world's population from distant times. MJ draw on these earlier efforts, which include Clark (1968) and Durand (1974). (Online Appendix Table A.1 summarizes the leading examples.) Caldwell and Schindlmayr (2002) discuss the intellectual history of these research projects, stressing their skepticism about the apparent consensus in the figures. MJ's effort differs from their predecessors in one important respect: the earlier estimates pertain to large regions or continents. MJ usually report populations for the areas that correspond to modern nation-states.

How did MJ derive population estimates from before, as they say, there was anything on which to base such estimates? Reading their descriptions and examining the figures suggests four overlapping approaches. In some cases, they state explicitly their reliance on one of these approaches, but more often, their method only reveals itself in the estimates. First, they start with the earliest official census and work backward. What Clark (1968, p. 61) calls "jobbing back" can yield good population estimates given the right raw materials and technique. The population of a country in 1500 equals its 1600 population minus deaths and net emigrants, plus births in the period 1500–1600. Wrigley and Schofield (1981, chapter 7)

⁵ D'Alpoim Guedes et al. (2013, pp. 73–75) are especially critical of MJ's estimates for the western hemisphere.

⁶ Clark (2009, p. 1160) asserts that for the years prior to 1820, MJ "is largely a work of imagination." Austin is referring to Nunn (2008). MJ understood their work would not please academics: "Even cautious users may well find this a boring book; academics are certain to find it irritating as well. There are many countries whose populations are not known with any certainty today. When we start giving figures for the dim and distant past, better-qualified hackles than ours are going to rise" (McEvedy and Jones 1978, p. 10).

offer an example of this approach. They start with the reliable census of England and Wales for 1841 and work back in time using estimates of births and deaths, along with more speculative estimates of net migration, to produce annual populations back to 1541. The challenge for earlier periods is that we rarely have anything like good counts of births and deaths, much less migrants, and the effort demands attention to complex sources. Creating the vital events series was the heart of Wrigley and Schofield's project.⁷

Austin (2008, p. 1002) stresses that momentous historical events such as the rise of the Atlantic slave trade greatly complicate such efforts. Few areas of the globe have been entirely spared these destabilizing episodes. Sometimes we even lack the equivalent of a reliable end-period enumeration, such as the 1841 census for England and Wales. Recent efforts to improve historical African population counts provide better-reasoned figures than MJ's for that continent, but run into a source problem. The twentieth-century colonial censuses that form their end-period figure are themselves not terribly reliable. In addition, for Africa, we lack the sources that would allow us to estimate the earlier population increases needed for useful "jobbing back."⁸

A hint comes from the suspiciously round progression of population figures for single countries.⁹ Table 1 shows the overall patterns; in many cases, MJ apparently devised a population estimate after deciding on a round figure for percentage growth. The many commonalities across countries are implausible. Individual country histories drive home the problem. In MJ's reckoning, England's population grew by 750,000 between 1600 and 1650, and by another 750,000 in the next half-century (McEvedy and Jones 1978, p. 43). Austria added 250,000 people every 50 years between 1650 and 1800 (pp. 88–92). Thailand added 250,000 people in both the sixteenth and seventeenth centuries (p. 193). Burma's population growth during the same period was 500,000 per century.

⁷ Scholars have questioned Wrigley and Schofield's estimates in many ways, but these criticisms rely on the fact that the original authors had substantive information and clear methods with which others could disagree. Deng (2004) argues that MJ ignored the possibility of using reasonable Chinese sources to do something similar for that country.

⁸ Frankema and Jerven (2014) describe these efforts and their proposal to improve them. All concerned with this research stress that Africa's population growth rates before the twentieth century have to be guessed by analogy to growth rates in other parts of the world. Gervais and Mandé (2007) document the many difficulties that afflicted colonial efforts to count populations in France's African empire.

⁹ Unless otherwise stated, calculations from MJ reported in this paper rely on a dataset provided by James Fenske. MJ report some areas as regions rather than countries, a practice Fenske retained. I refer to this data as the "MJ database." MJ's figure for France has an obvious misprint that I have corrected; their graph implies that the population in 1600 was 18.5 million, not 10.5 million.

TABLE 1
SUMMARY OF INTER-PERIOD PERCENTAGE CHANGES IN MJ

Period	Number of Countries in Dataset	Percentage of Countries for Which Change in Population Is "Round"	Most common values					
			Most Common		Second Most Common		Third Most Common	
			Value	Percent	Value	Percent	Value	Percent
1–1000	82	63	100	35	0	9	25, 50 & 66	5
1000–1500	96	53	100	20	0	12	50	8
1500–1600	79	30	25	21	0	16	50	8
1600–1700	79	39	0	28	20		25 & 50	8
1700–1800	99	42	0	14	50	11	20	8
1800–1900	115	27	100	10	66	5	50 & 150	4

Notes: "Round" means the inter-period percentage change is evenly divisible by 50. I computed the percentage changes and then rounded them to the nearest integer value (e.g., 33.2 becomes 33 percent). This procedure has no effect until the last period.

The table shows that, for example, of the 79 countries for which MJ reported data in 1500 and 1600, 30 percent have an implied percentage change in population that is a round figure. The modal change for that period was 25 percent; MJ think 16 percent of countries had that modal growth rate. The second most common figure is 0 percent and the third is 50 percent.

Source: Computed from the MJ database.

Second, MJ apparently wanted their estimates to reflect their view that until the late medieval period, population grew at a constant rate. In disagreeing with an earlier author on the right total world population for the year 1000, MJ note that “our figure for AD 1, being 100m below the agreed figure for AD 1000, fits better on the sort of exponentially rising curve that everyone agrees best describes mankind’s population growth” (p. 354). As the quotation implies, MJ also worried about consistency between theirs and earlier estimates. Caldwell and Schindlmayr (2002, p. 199) call this “an example of a dangerous circularity,” while Biraben dismisses the MJ data after noting this fact.¹⁰

Third, in the face of ignorance, MJ felt comfortable assigning identical growth rates to places they thought were similar. This practice doubtless underlies much of what we see in Table 1. For 35 percent of countries, MJ assign the same figure to population growth between the years 1 and 1000.

Finally, especially before 1500, MJ tended to reason from the nature of an economy and the population they thought it could support. They are rarely explicit about this tactic, but it shows through remarks such as “likely population densities” in the passage about Poland quoted earlier. To the extent that they estimate population in this way, MJ’s figures reflect not the population of a particular country at a point in time, but their views about the population density consistent with the kind of economy MJ thought the country had. Since they do not claim any serious knowledge of the economy or of the number of people it can support, the basis for this reasoning is unclear.¹¹

Maddison

Several of the articles discussed later rely in part on estimates reported by the late Angus Maddison. Maddison famously constructed, updated, and used a database that offered estimates of population and GDP/capita for most of the world’s countries, again, in some versions, going as far back as the year 1 CE.¹² For the last major revision of his estimates,

¹⁰ Deng (2004, note 3) also stresses this problem. Biraben (1979, p. 13): “Most authors, in fact, when they describe the population of the world through the ages, evade the difficulties and uncertainties relating to ancient periods by making the hypothesis of a more or less regular exponential growth” (my translation).

¹¹ Caldwell and Schindlmayr (2002, p. 199) refer to MJ as “geographers.” It is true that McEvedy (1930–2005) was famous for a series of atlases intended for general audiences, but he was actually a professional psychiatrist. He wrote seven works for the Penguin Historical Atlas series alone, and three more contributions to the Atlas of World History with his wife, Sarah McEvedy. The *Atlas of World Population History* is one of four additional works intended for a broad audience; see https://en.wikipedia.org/wiki/Colin_McEvedy (accessed 14 January 2022).

¹² The effort continues in the form of the Maddison Historical Statistics project. See <https://www.rug.nl/ggdc/historicaldevelopment/maddison/>.

Maddison says of his population data: “The following detailed estimates for 1500 onwards rely heavily on monographic country studies for the major countries. To fill holes in my dataset I draw on McEvedy and Jones (1978). For the preceding millennium and a half, I use their work extensively” (Maddison 2001, p. 230). Maddison adds that he relies on MJ rather than earlier accounts because MJ are “the most detailed and best documented” (p. 230).

Thus, for many places before 1500, Maddison’s database just reproduces MJ’s figures. This is not always the case, however; Maddison was able to incorporate the fruits of research published between 1978 and his own publication. This led to some substantive revisions, but those revisions reflect the research literature’s emphasis. He updated 23 percent of MJ’s observations for the year 1000, for example, and 40 percent of the observations for 1500. The majority of Maddison’s changes for the year 1000 were in non-European countries (eight of nine countries that changed were outside Europe). For 1500, this pattern changes; 10 of 16 changes are for European countries, and in 1700, 8 of 12 are for Europe. These changes reflect contributions from the research literature.

Some individual changes are much larger, however. Maddison added 50 percent to Mexico’s population for the year 1000, and he doubled Peru’s population in that same year (Maddison 2001, table B-5, p. 235). He increased the population of the territory that would become the United States by 125 percent for the year 1500. For later periods, especially in the twentieth century, Maddison revises the MJ estimates more comprehensively. In 1850, 84 percent of Maddison’s 51 observations have values different from MJ’s, although the average absolute difference (3 percent) is smaller than for earlier years.¹³

MEASUREMENT ERROR AND ROUNDING

Relative to “perfect” data for every country in the world, how far wrong will MJ take us? It is worth reviewing some general consequences of measurement error for the kinds of linear models that most researchers use.¹⁴ Denote the true population of country i in year t as \dot{P}_{it} . The MJ estimate is P_{it} . The difference between MJ’s estimate and the true population is the measurement error ε_{it} , such that $P_{it} = \dot{P}_{it} + \varepsilon_{it}$. Classical measurement

¹³ The numbers in the text reflect only countries that appear in both the MJ and Maddison datasets. Maddison did not hazard guesses for many of the countries that appear in MJ. I adjusted the figures so MJ’s rounding would not create a false difference with Maddison. Maddison (2007) offers further revisions and extends some of the estimates back to ancient times, but most of these revisions pertain to the twentieth century.

¹⁴ Standard textbook treatments include Greene (2018, pp. 102–3, 281–88) as well as Wooldridge (2010, pp. 78–82).

error is the special cases where ε_{it} is additive and uncorrelated with \dot{P}_{it} . We have two general implications. First, classical measurement error in the dependent variable alone does not bias estimates. The ε_{it} are swept into the regression error term, and the only consequence is some efficiency loss. Second, measurement error in *any* regressor implies bias in *all* of the estimates.

Consider the following regression:

$$Y_{it} = \alpha + \beta X_{it} + \gamma P_{it} + \mu_{it}, \tag{1}$$

where P_{it} is the mis-measured variable. While I write Equation (1) for a panel framework, that is not necessary for what follows. Classical measurement error in P_{it} implies that the estimate for γ will be smaller in absolute value than it would be if we could use \dot{P}_{it} instead. The estimate is attenuated. The estimate for β will also be biased in ways we cannot ordinarily sign. The problem arises from the correlation between the measurement error ε_{it} and the regression error term μ_{it} , which is why some researchers employ instrumental-variable techniques in using the MJ data as a regressor. The fixed-effects estimator does not necessarily yield unbiased estimates in the presence of even classical measurement error. Fixed effects only “deals with” measurement error if the errors in P_{it} are, for each country i , the same for all years. In that case, the measurement error becomes part of the estimated country fixed effects (Deaton 1997, pp. 108–110).¹⁵

Classical measurement error in the dependent variable ordinarily does not bias regression estimates because the measurement error is added to the regression disturbance term. This result requires that the measurement error be additive: $P_{it} = \dot{P}_{it} + \varepsilon_{it}$. One common case of non-additive measurement error appears when the dependent variable is the ratio of two variables and the denominator is measured with error. Consider a common example: an urbanization figure is formed as the number of people living in cities divided by MJ’s population estimate. Rewriting Equation (1),

$$C_{it} / P_{it} = \alpha + \beta X_{it} + \mu_{it}, \tag{2}$$

where C_{it} is the urban population. Using MJ’s population estimate implies that the denominator is the true population plus measurement error, $P_{it} = \dot{P}_{it} + \varepsilon_{it}$. Substituting and re-arranging, we have:

$$C_{it} = \alpha(\dot{P}_{it} + \varepsilon_{it}) + \beta X_{it} (\dot{P}_{it} + \varepsilon_{it}) + \mu_{it} (\dot{P}_{it} + \varepsilon_{it}) \tag{3}$$

¹⁵ If the errors are the same for all countries in a year, then they become part of the year fixed effects.

The ratio in the original dependent variable makes the measurement error multiplicative and causes bias in estimates of β . More generally, if measurement error is not classical, then we need to model the error. Hyslop and Imbens (2001) discuss several cases, including one where measurement error in a regressor leads to overestimates of that coefficient instead of the attenuation we expect with classical measurement error.

What does this mean for econometric studies that used mis-measured population estimates? If we maintain the assumption that the measurement error is classical, we can say two things. When population is the dependent variable, the estimates may be less efficient, but there should be no bias due to measurement error alone. If population is a regressor, on the other hand, then the estimate for population will be attenuated. Additionally, the other estimates in this case will be biased and inconsistent. Thus, using population as a “control” can lead to bias even for variables not thought to suffer from measurement error. If the error is not classical, on the other hand, then we cannot say much without modeling the measurement error.

While economists tend to assume that measurement error is always classical, in this case, we know that this is not the case. MJ state that they have rounded their estimates in ways that make the measurement error depend on the true value. This rounding applies to every country and every period, but the rounding rule depends on the population size. This means the measurement error depends on population size:

All figures are rounded on the following system: below one million to the nearest .1 million, between one and 10 millions to the nearest .25 million, between 10 and 20 million to the nearest .5 million and between 20 and 100 millions to the nearest million. Above 100 million the rounding is to the nearest 5 million, above a billion... to the nearest 25 million. (McEvedy and Jones 1978, p. 9)

Thus, MJ tell us that they create measurement error that is larger for larger populations. We cannot know precisely the implications of MJ’s rounding rules. We can, however, simulate the “true” populations to get a feel for how much trouble the rounding can cause. I use a Monte Carlo exercise to simulate the rounded-off portion of each population estimate. Adding that rounded portion to MJ’s reported numbers yields a simulated “true” population. We can then ask whether that simulated “true” population is correlated with the error caused by rounding. This exercise can only address the measurement error caused by rounding; the other flaws remain. Table 2 shows the result: the rounding induces a high degree of correlation between the measurement error and the population. This result holds for four different assumed functional forms

TABLE 2
SIMULATING UNROUNDED FIGURES IN MJ'S DATA

	Year		
	1000	1500	1800
Panel A: The Percent of Experiments in Which $p < .10$			
Uniform	72.2	83.4	86.6
Normal	86.7	94.3	93.4
Beta	80.2	86.9	89.5
1- beta	79.2	84.8	89.9
Number of valid observations in MJ data that year	97	101	118
Panel B: Number of Countries in MJ's Rounding Categories			
Less than 1 million	55	44	46
1–10 million	40	50	61
10–20 million	0	5	5
20–100 million	2	0	4
100 million to 1 billion	0	2	2

Notes: The upper panel reports the results of 1000 Monte Carlo draws that assume the stated distributional form for the population value MJ rounded-off. Each experiment computes the correlation between the simulated “true” value and the simulated error induced by rounding. The p-values are for the null hypothesis that the correlation is zero. Small p-values for the correlations indicate violation of the classical measurement-error assumption. See text and Online Appendix Section 2 for details on computation. The beta distribution assumes parameters 1 and .5. The beta distribution is asymmetric; “1- beta” places the thicker part of the density on the left-hand side instead of the right-hand side. Panel B reports the number of countries affected by MJ's different rounding rules. No countries in MJ have populations larger than 1 billion in these years, so their last rounding category does not appear in the table.

Sources: MJ data and own calculations.

for the rounding error, including two that are asymmetric in different ways. The correlation stems from MJ's different rules for different size categories.

Online Appendix Section 2 reports details of this simulation along with two additional assessments of the importance of this rounding. The first uses the populations of the 50 United States for the period 1900–1970. The second uses the populations of countries around the world for the period 1960–2020. In both cases, I apply MJ's rounding rules and examine the correlation between the true values and the errors created by rounding. Rounding for the U.S. states does not consistently imply correlation, while for the countries of the world, the correlation between the true value and the measurement error is considerable. The countries dataset is the closer analogy to MJ because the countries span the entire range of their rounding rules.

MJ's rounding procedure creates a distinct problem when a country's population crosses one of the thresholds implied by their rounding

rule. Portugal, they report, had a population of 900,000 in 1400 and 1.25 million in 1500 (p. 103). These figures imply that Portugal's population increased by 350,000 people, or 39 percent, in those 100 years. Taking the rounding into account, however, implies upper and lower bounds for the population estimate in both 1400 and 1500. The true increase could be as small as 19 percent or as great as 58 percent.

The non-classical nature of the measurement error in MJ poses a serious problem for any estimates that rely on it. We can evaluate earlier, published work under the assumption of classical measurement error, and that is not a bad place to start. But MJ's rounding applies to every country and period in their data, which means that none of the standard intuitions based on classical measurement error really apply.

CIRCULARITY

Many economists who use MJ's figures think of population (or a derivative such as population density) as a proxy for an economic aggregate such as output. Critics such as Caldwell and Schindlmayr (2002) and Austin (2008) note MJ often use ideas about the economy to derive an estimate of population size, thus making the population estimates a poor proxy for an economic aggregate. This is especially true in places and times for which the population data are thin. As noted, MJ defend an estimate for medieval Poland by referring to "likely population densities." In a more explicit example, MJ discuss agricultural conditions in a region that comprises the modern states of Columbia, Venezuela, and the Guyanas to defend their assumption that until 1500, Colombia always accounted for 2/3 of the region's population (McEvedy and Jones 1978, p. 302).

Austin stresses that this approach makes their estimates hostage to ideas about an economy and economic change. It is a particular problem for Africa because we know relatively little about that continent's economic history. Maddison (2001, p. 238), for example, adopts MJ's estimates for Africa in preference to earlier alternatives because MJ "assumed a more dynamic growth process." That is, Maddison preferred MJ's *population* estimates because he agreed with their assessment of the African *economy*. Neither Maddison nor MJ offer independent evidence about the African economy. To the extent MJ assigned population estimates based on their perceptions of economic performance, a regression using population as a proxy for growth tells us more about MJ than about economic growth.

SOFT CLONES

Researchers who use MJ's data treat them as if they imply independent observations; put differently, if there are N countries listed for a given year, this reflects N pieces of information. This is not always true, for two distinct reasons that I will call "soft" and "hard" clones. MJ themselves create the soft clones. Frankly admitting that they lack meaningful data, they assign to some countries the population dynamics of countries they think are similar. Sometimes they make this approach explicit. After concluding that Afghanistan has no useful population data before the twentieth century, MJ say that "Perhaps the best approach is to compare Afghanistan with Iran" (McEvedy and Jones 1978, p. 156). What they did, in fact, was to assume that Afghanistan had half the population of Iran in every year before 1900. The measurement error for Afghanistan thus has two sources. The Afghan numbers share any measurement error in the figures for Iran, and they also suffer from the error implied by any deviation of Afghanistan's true population dynamics from Iran's.

MJ includes many soft clones. Kenya and Uganda, for example, had identical populations through 1800, although the text does not say why. In some cases, they appeal to the idea that neighboring countries should have similar population growth rates: "... the fact that population doubled in most European countries between A.D. 1000 and 1300 can be taken as strong evidence for it doing so in other European countries for which direct evidence is lacking" (p. 11). Thus, in their reckoning, Poland, Hungary, and Czechoslovakia each grew 20 percent between 1000 and 1100. In the fifteenth century, European Russia and China each grew by one-third. As late as 1600–1700, Romania and Austria each grew by 11.11 percent. Soft clones probably underlie the patterns we see in Table 1.

HARD CLONES

A final problem reflects both MJ's estimates and the way some economists have used them. MJ report many populations for regions rather than modern countries. Some economists create country-level populations out of the regions by allocating the regional population among the constituent modern nation-states. I will call the resulting countries "hard clones." In the cross-section, these clones differ in size within the region. By construction, however, in the time series, all members of a clone group share the population growth rate MJ assigned to the region. Hard clones account for an especially large portion of the African country-level observations, but they appear in other parts of the world, as well. In Nunn

and Qian (2011), hard clones account for 76 percent of the observations in Africa, 36 percent in Europe, and 41 percent in Asia. In Ashraf and Galor (2011), the clones are similar for these continents; three-quarters of their Western Hemisphere countries are clones.¹⁶

The literature includes two different ways to create countries out of regions. Nunn (2008, p. 170) assumes that the relative sizes of the populations within each region are the same as reported for 1950. Nunn and Qian (2011) do not say explicitly how they disaggregated the regions, but for most countries, their population figures are similar to Nunn's, so the approach is probably similar.¹⁷ Ashraf and Galor (2011, 2013) disaggregate the regions by assuming that each country within a region has the same population density in each year. In general, the resulting "country" populations created by the two methods differ in the cross-section; Nunn and Qian's Nigeria in 1500 is not the same size as Ashraf and Galor's. Yet both cloning methods imply that Nigeria has the same growth rate between any two years. This growth rate is simply the rate implicit in the region from which Nigeria is cloned.

Hard cloning adds further error to MJ's guesses; how much is something we cannot say precisely because we do not know the true populations of the clones in those years. We can, however, study the implications of these two methods in contexts where we have the equivalent of valid country-level numbers. For the years 1900–1970, I constructed a panel from the population of the 50 United States as reported in the decennial census. I then aggregate the state populations into four standard regions. The state populations (which we know) are analogous to the unknown country populations that hard cloning attempts to recover. The U.S. regions are like the regions in the MJ book. I apply both the Nunn-Qian and Ashraf-Galor methods to estimating the population of each state in the period 1900–1960, as if all I knew was the population of each state in 1970 (for Nunn-Qian) and the state area and regional population in each year 1900–1970 (Ashraf-Galor). Table 3 summarizes the errors these methods produce. In most years, the Nunn-Qian approach produces smaller errors than Ashraf-Galor, although those errors are still

¹⁶ Online Appendix A.4 provides detail on the geographic distribution of the hard clones. Guinnane (2021, figure 2) is MJ's "Equatoria, Zaire, and Angola" (p. 249), which includes ten states. The region MJ call "West Africa" (McEvedy and Jones 1978, p. 243) creates 12 observations in the Nunn-Qian and Ashraf-Galor datasets.

¹⁷ "For groups of some smaller countries, population data are only disaggregated to a regional level" (Nunn 2008, p. 170). The aggregation also affects some of the largest countries, including Nigeria and Congo in Africa and Ukraine in Europe. Nunn and Qian (2011, p. 644) say only that "country-level population data are from McEvedy and Jones (1978)"; the online appendix to that paper does not discuss this issue, nor does their replication code show how the hard clones were created.

TABLE 3
THE ERRORS CREATED BY APPLYING HARD CLONE METHODOLOGY
TO U.S. STATES, 1900–1970

Year	Median		25th percentile		75th percentile	
	NQ	AG	NQ	AG	NQ	AG
1900	20.05	43.80	-7.30	-29.26	38.21	63.41
1910	18.17	25.44	-8.09	-39.65	34.31	64.18
1920	16.85	4.95	-6.89	-150.56	31.18	40.95
1930	9.74	0.74	-10.93	-176.84	25.66	32.81
1940	9.59	0.68	-9.79	-148.66	21.04	45.71
1950	5.88	-0.15	-5.35	-153.68	15.31	45.67
1960	1.69	-9.19	-2.33	-203.12	5.91	41.38
1970	0.00	-5.62	0.00	-188.19	0.00	39.05

Notes: All figures are the error as a percentage of the actual state population in that year. The error is defined as the actual population minus the population implied by the method in question. There are 50 states in each year (the tables consider territories that became states later as states). “NQ” (Nunn-Qian) assumes that the relative population sizes within a region in 1970 were true in all previous years. For this method, 1970 is accurate by construction. “AG” (Ashraf-Galor) assumes that every state within a region has the same population density each year. The calculations assume four regions: Northeast, Midwest, South, and West. See text and the Online Appendix for details.

Source: Computed from MJ database.

large. Only in 1960 did the median error from the Nunn-Qian approach fall below 5 percent of the actual state population. The Ashraf-Galor approach produces a smaller median error in the early years, but the variance of the errors using this method is large. The two types of error are not highly correlated in the cross-section.¹⁸

What does this exercise tell us about disaggregating MJ’s regions? The Nunn-Qian method assumes that population growth rates within each region are similar over time. The method goes wrong for regions with a state like California, which experienced especially rapid growth in the twentieth century. In 1900, California’s population accounted for 34 percent of the “West” region; in 1970, it was 57 percent. This is probably why the Nunn-Qian approach improves monotonically better over time; the 1970s weights better approximate the population distribution in 1960 than in 1900. It also illustrates the danger of their research, which starts with MJ’s estimates for 1000. Between the years 1000 and 1950, there was plenty of opportunity for the countries within a region to grow at different rates, producing a version of the problem noted for California.¹⁹

¹⁸ The exercise includes 50 states every year by treating territories such as Alaska or Hawaii as states before their actual statehood. See Online Appendix Section 3 for more detail.

¹⁹ In a personal communication, Quamrul Ashraf noted this feature of the Nunn-Qian approach as a reason to prefer Ashraf-Galor.

Ashraf and Galor's approach, on the other hand, requires that the population densities for countries within a region be identical. This approach fares poorly in the U.S. case because of the unequal population densities within some U.S. regions; the "Midwest" region, for example, includes states like Ohio (204 persons per square mile in 1900) as well as states like North Dakota (nine persons per square mile). The assumption is unlikely to hold at any point in time, and the discrepancy between assumption and reality could change over time with the introduction of new crops or other changes that lead to uneven economic development within the region. The U.S. experience may not provide a strict analogy to the regions these two methods attempt to disaggregate, but this exercise highlights how far wrong things can go if strong assumptions do not hold. The U.S. case also highlights the questions that we would need to ask before disaggregating data in this way. Do we really know enough about the sub-regional patterns in the Sahel in 1000, for example, to divide up a regional population?

The hard clones play an especially important role in Africa. MJ report only 12 regions for Africa. Nunn (2008)'s Africa has 52 countries, while Nunn and Qian (2011)'s have 47.²⁰ Three-quarters of the African observations are thus clones. Given the difference in methods, we expect Nunn-Qian and Ashraf-Galor to assign different populations to the same country, but the differences can be huge. Online Appendix Figure 1 reports the distribution of the ratio of Ashraf-Galor's clones to Nunn-Qian's for the Old World in 1500. This figure illustrates the great range in values for a given place and time that result from cloning the MJ regions. In Africa, this ratio ranges from .192 (Malawi) to .61 (Nigeria) through South Africa (1.003) to Congo (2.208) and Côte d'Ivoire (2.887).²¹

The disaggregation problems account for only one of two different sources of measurement error for hard clones. The first comes from MJ itself; MJ's regional estimates are themselves noisy and rounded. Cloning assigns that noise to each of the country-level figures and adds additional error because we do not really know what the right allocations within a region should be. This additional disaggregation error is, by definition, negatively correlated for countries within a given MJ region and year. The cloned population estimates cannot be "correct." The implications for change over time, however, are the same: every clone from a given

²⁰ The difference in the number of observations reflects five small islands that Nunn (2008) treats as countries but that Nunn and Qian (2011) exclude. Alternative specifications in Nunn (2008) show that these five observations do not drive the results.

²¹ These calculations compare Nunn and Qian (2011) to Ashraf and Galor (2013), using the respective replication files for the year 1500. The differences are not confined to Africa; the ratio for Ukraine is .453 and 2.439 for Pakistan.

region *must* grow at the same rate, the growth rate MJ assigned to the region. Breaking these regions up into observations does not create more information. It just creates clones.

Econometric Implications of Hard Clones

We cannot re-estimate earlier models using correct population data because that is obviously not available. The best check on the implications of cloning would be to dispense with the disaggregation and re-estimate the models using the units MJ reported. Since that check would require redefinition of all the other variables as well, it lies beyond the scope of this paper. We can, however, show that relying on clones significantly affects the results of published research. The following discussion only considers two cases and focuses on this issue alone. Table 4 considers the baseline results from Nunn and Qian (2011), which studies the old question of whether the potato's introduction in the Old World caused population growth. The regressions use the population of all countries in the Old World at century intervals between 1000 and 1900, along with the years 1750 and 1850. The dependent variable is always population. Although they report and discuss other specifications, Nunn and Qian focus on models in which the regressor of interest is the interaction between an index of the fraction of a country's land that is suitable for potato cultivation and a dummy for the years 1750 and later. They regard this interaction as a proxy for the effect of the potato's actual introduction.²² Every specification includes year and country fixed effects. Some models have no additional controls; we focus on models that include the "baseline" controls.

Table 4, Column (1), reproduces the result from Nunn and Qian (2011, table IV, column (1)). As they stress, the interaction implies that the potato's introduction increased population sizes in the years 1750 and later. When we drop the African clones (Column (2)), however, the point estimate (and the average marginal effect) are no longer significantly different from zero. Dropping all clones (Column (3)) does not produce this effect; the problem appears to be the African clones. On the other hand, if we drop all of Africa (Column (4)), the point estimate and average marginal effect (AME) become even smaller. Given that about three-quarters of the African observations are clones, it is difficult to know whether Africa in general does not fit the story or if there is something particular to African clones.

²² Online Appendix Section 4 discusses the "full-flexible" results Nunn and Qian report in their Table II.

TABLE 4
THE ECONOMETRIC CONSEQUENCES OF HARD CLONES

Cloned Using:	(1) Nunn-Qian	(2) Nunn-Qian	(3) Nunn-Qian	(4) Nunn-Qian	(5) Ashraf-Galor	(6) Ashraf-Galor	(7) Ashraf-Galor
Potato × post 1750	0.0325 (0.0116)	0.0296 (0.0172)	0.0575 (0.0209)	0.0207 (0.0184)	0.0330 (0.0116)	0.0312 (0.0171)	0.0223 (0.0183)
Constant	12.64 (0.0324)	12.84 (0.0433)	13.29 (0.0563)	12.77 (0.0448)	12.66 (0.0322)	12.82 (0.0429)	12.74 (0.0444)
Observations	1,552	1,108	712	964	1,552	1,108	964
R-squared	0.990	0.990	0.985	0.991	0.991	0.991	0.992
Sample	All	-Afr clones	No clones	No Africa	All	-Afr clones	No Africa
AME	.003	.003	.006	.002	.003	.003	.002
SE of AME	.001	.002	.002	.002	.001	.002	.002

Notes: The dependent variable is population. The regressions have fixed effects for time period and country. The standard errors are clustered by country. The models in (2) and (6) drop all clones on the continent of Africa. All specifications include the “baseline” controls. See Nunn and Qian (2011, table IV). In Columns (1)–(4), the hard clones are computed using the Nunn-Qian approach; in (5)–(7), using the Ashraf-Galor approach. “AME” is the average marginal effect for the potato/post-1750 interaction.

Source: Computed from the replication files for Nunn and Qian (2011).

The next three specifications repeat (2)–(4) using Nunn and Qian’s data but use the Ashraf-Galor definition of clones. The results differ somewhat, but the overall message is the same. In this model, it does not matter how we construct the clones; dropping Africa’s clones, all clones, or all of Africa has the same effect as with Nunn and Qian’s definition. Table 4 holds two lessons. First, the Nunn-Qian result depends critically on the inclusion of Africa or its clones. This may not hold for all of Nunn and Qian’s specifications, although Online Appendix Section 4 demonstrates the same problem in their fully flexible approach. Second, and more generally, in this example, it does not much matter how we construct the hard clones. This follows from including country-level fixed effects. Since identification comes from within-country change and since the clones, however constructed, all have the same growth rates as the regions from which they are disaggregated, in this type of model, the error created by cloning does not depend on how the clones are constructed.²³

Fixed-effects models cannot cure measurement error in general, as I stressed earlier. Even the countries that are not clones in the Nunn-Qian or Ashraf-Galor datasets have rounding error plus the measurement error inherent in MJ’s guesses. This replication exercise makes a narrow and specific point. I have shown that first, Nunn and Qian’s results depend on including African observations that are really clones, and second, there is no important difference between the additional measurement error created by two different ways of creating hard clones.²⁴ Surely those who will rely on MJ in the future should at least dispense with the clones and use as their units of analysis the regions that appear in the population data.

HOW ECONOMISTS USE MJ

To obtain a more specific idea of how economists use these data, I examined every paper that cites MJ that was published in one of the “Top 5” economics journals through 2020.²⁵ I set aside many of these papers for the rest of this discussion. This list includes a few articles that cite MJ

²³ Researchers typically cluster their standard errors at the level of a country. Extending the clustering to include all members of a clone group for all years increases the reported standard errors. In this sense, the standard errors reported in articles such as Nunn and Qian (2011) or Ashraf and Galor (2011) are lower bounds. Clustering, of course, does not affect the point estimates.

²⁴ Online Appendix Section 4 also reports a similar exercise for Ashraf and Galor (2011). In that case, the regressor of interest remains significant when I drop the clones, but the magnitude of the point estimates changes dramatically.

²⁵ These are *The American Economic Review*, the *Quarterly Journal of Economics*, *The Review of Economic Studies*, and the *Journal of Political Economy*. *Econometrica* did not publish any papers meeting this criterion.

but do not use the data in econometric exercises. Shiue and Keller (2007, p. 1194), for example, cite MJ and other authorities as implying that their two regions, China and Europe, had similar populations at the end of the eighteenth century. Rogers (1994, p. 467) cites MJ to defend the assumption that long-term population growth rates were nearly zero until relatively recently. This usage seems consistent with the spirit in which MJ offer their estimates. I also set aside papers that only use MJ's estimates for 1900 and later. By that date, the information MJ reports comes almost entirely from reasonable census reports (although they round even these figures). This includes articles such as Acemoğlu, Johnson, and Robinson (2001).²⁶

A first question pertains to dates; the MJ data are more suspect in earlier periods. The year 1500 does not form a magical dividing line, but it is the earliest year for which we have anything like reliable estimates for populations of even most European countries, which tend to have the best-founded estimates. Several articles depend in a serious way on MJ's population estimates from before 1500. Ashraf and Galor (2011) report econometric results that depend critically on population data from the years 1, 1000, and 1500. Population is the variable of interest in Nunn and Qian (2011), which starts with the year 1000. Nunn (2008) uses the 1400 estimates alone.²⁷ Several other papers also rely on data from 1500–1800.

A second issue is whether MJ's population figures form the dependent variable or a regressor. Many articles use population as the dependent variable, where it does least harm under the assumption of classical measurement error. These include Ashraf and Galor (2011, 2013) and Nunn and Qian (2011).²⁸ In others, the MJ data scale the dependent variable. As noted previously, this means the measurement error in the dependent variable is not classical, and the estimates are biased in unpredictable ways. Acemoğlu, Johnson, and Robinson (2005)'s urbanization regressions are an example.

Some articles, however, create regressors from MJ's estimates. This list includes Iyigun (2008) as well as Gennaioli and Voth (2015). Iyigun (2008) studies whether military pressure from the Ottoman Empire helped reduce conflict among European states in the early-modern period. The

²⁶ I also set aside two intermediate cases. Putterman and Weil (2010, note 11) rely on MJ's year 1500 estimates for some of their analysis, but they first aggregate the data to 11 large regions "because population data for 1500 are very noisy, particularly at the country level." Voigtländer and Voth (2013) rely on MJ's figures for Europe in the period 1000–1700, but do not report econometric estimates.

²⁷ "Rely on" in the sense that these years are part of the sample used to estimate the primary regression results. Nunn and Qian (2011) do not need data from the western hemisphere.

²⁸ Acemoğlu, Johnson, and Robinson (2005) use MJ's population figures to weight their regressions.

econometric models rely on annual observations for the period 1450–1700. The dependent variables measure intra-European conflict. The controls include measures of Ottoman military pressure as well as the populations of Europe and, in some specifications, the Ottoman Empire's. Iyigun (2008, p. 1476) describes the population data as a proxy for economic "size and strength." The estimated effect for European population size is imprecisely estimated in most specifications, while the Ottoman population variable is more precisely estimated but switches signs, depending on the dependent variable. The point estimates for both population variables *must* be attenuated if this is classical measurement error, so we cannot really say whether Europe became more peaceful simply because of economic growth, nor can we assess the implications of Ottoman economic conditions for European conflict. Moreover, the estimates for his main variable of interest, the extent of Ottoman military incursions into Europe, may be biased because of the measurement error in population.

Gennaioli and Voth (2015, table 3) address a related question, and their population figures cause similar trouble. They study the determinants of battle success in early-modern European conflicts. The authors set this up as a horse race between fiscal strength on the one hand and population size on the other. Greater fiscal strength allows a state to pay more mercenaries and support more allies. Population size could matter in early-modern war because larger populations make it easier to field larger armies. In most specifications, the fiscal variable has a positive and significant effect on battlefield success, while the relative populations of the two combatants have almost none. They conclude that "Differences in population size do not have a systematic effect on the chance of battlefield success" (Gennaioli and Voth 2015, p. 1430). This result could reflect nothing more than the measurement error in MJ's estimates.

A third issue pertains to how the authors confront the possibility of measurement error in the population data. Acemoğlu, Johnson, and Robinson (2002) and Acemoğlu et al. (2008) explicitly discuss measurement error and use IV methods to contend with measurement error in regressors. Others take a different approach. Ashraf and Galor (2011, p. 2011) claim:

The most comprehensive worldwide cross-country historical estimates of population and income per capita since the year 1 CE have been assembled by Colin McEvedy and Richard Jones (1978) and Angus Maddison (2003), respectively. Indeed, despite inherent problems of measurement associated with historical data, these sources remain unparalleled in providing comparable estimates across countries in the last 2,000 years and have, therefore, widely been regarded as standard sources for such data in the long-run growth literature.

They do not argue that MJ's data meet any particular standard. Rather, they know of nothing better (it is "unparalleled") and everyone else uses it (it is the "standard source in the long-run growth literature").

Nunn and Qian (2011, p. 616) address measurement error more explicitly, but their discussion consists of general statements that are not relevant to the MJ data:

Accuracy is an obvious concern for historical data that span such a long time horizon and broad cross-section. However, classical measurement error in our outcome variables will not bias our regression estimates. Similarly, any systematic measurement error that varies by time-period or by country is captured by the country and year fixed effects, which are included in all specifications.

Population is their dependent variable, so they are correct that if the measurement error is classical, it does not bias their results. They provide no reason to think this is true, and, as noted, MJ say it is not true. The second statement about fixed effects is equally true but irrelevant to the case. Neither of these extreme assumptions is likely. Nor can they be true simultaneously.

CONCLUSIONS

We know the population of the United States in 2020 to a high degree of accuracy. For historical episodes, we might not always do as well, but we can definitely do better than MJ. Palma, Reis, and Zhang (2020), for example, provide improved estimates for Portugal (1527–1850) by combining two sets of historical estimates with some judicious reasoning. Similar approaches may be possible for other times and places. Federico and Tena-Junguito (2022) provide considerable improvement over MJ for the period since 1800 by making better use of published data. Earlier periods may require different approaches. Refining the estimates for Poland in 1400, for example, may not just require consulting more published works, but also original research using, for example, essentially archeological techniques. But it can be done. It would not be useful to assert that because we cannot know the population of Poland in 1400 with the same accuracy as we can in 2020, there is no point in using historical population counts. The opposite extreme is more common and pernicious: many economists take the view that the accuracy of historical data does not matter because it cannot be as precise as modern reports.

We can do, and have done, better. Consider one example. MJ's estimates imply that the population of England and Wales grew at an average annual rate of .32 percent in the period 1600–1650. Wrigley and Schofield's

figures put that rate at .5 percent. For the period 1650–1700, the estimated growth rates are .26 in MJ and -0.07 in Wrigley and Schofield; for 1700–1750, they are .10 and .26. The differences are substantial. MJ missed the population stagnation of the second half of the seventeenth century and significantly understated the population growth of the first half of the eighteenth century. Population figures directly underlie, any statement about per-capita GDP or its growth rate and are thus central to understanding the Industrial Revolution. We can do much better than MJ's guesses for many countries, especially in the period since 1500 (e.g., Vos 2014, pp. 366–69).

Econometric estimates that rely on MJ form a particular literature that takes a “cross-country regression” approach to economic growth, political economy, and related questions. Economists differ on the usefulness of the general research strategy, and those who favor such studies may insist that some data are better than none. Even those who take this view, however, should be aware of the pitfalls of the source and the way some use it. As I have noted, Acemoglu and his co-authors tend to use MJ as carefully as one can. Others have compounded MJ's weaknesses by trying to create information that is not in the source. Some of the research discussed here appeals to the idea that classical measurement error does not cause bias in linear models when the measurement error affects only the dependent variable. This observation is mathematically true but not relevant to the MJ data.

This paper documents a series of problems in a published source that underpins many articles published in the leading general-interest economics journals. Publication in these outlets has strong professional rewards and conveys signals. One signal is that if everyone does something inappropriate, then it is fine. A second signal discourages the original work necessary to improve the basis of our knowledge. The researchers who did the groundwork on which MJ is based understood themselves as contributing to a broader literature in the social sciences. Their contributions were rewarded within their own niches. The same applies to all of the effort that went into constructing the considerable information on historical economies that Maddison summarizes. To the extent the profession signals a lack of interest in such work, it is unlikely we will ever learn more about, for example, the population of Poland in 1400.

This discussion holds a simpler lesson. Many economists today download a dataset and merge it into other datasets without consulting the original sources. Examining MJ's book is instructive. The introduction explains the problem of non-classical measurement error. A look at the graphs (such as my Figure 1) would lead most to treat the data with considerable caution.

Anyone looking at MJ's maps for Africa should wonder why their Africa has so many countries in the data.

REFERENCES

- Acemoglu, Daron, and Simon Johnson. "Unbundling Institutions." *Journal of Political Economy* 113, no. 5 (2005): 949–95.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91, no. 5 (2001): 1369–401.
- . "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117, no. 4 (2002): 1231–94.
- . "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review* 95, no. 3 (2005): 546–79.
- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. "Income and Democracy." *American Economic Review* 98, no. 3 (2008): 808–42.
- Ashraf, Quamrul, and Oded Galor. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101, no. 5 (2011): 2003–41.
- . "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* 103, no. 1 (2013): 1–46.
- Austin, Gareth. "The 'Reversal of Fortune' Thesis and the Compression of History: Perspectives from African and Comparative Economic History." *Journal of International Development: The Journal of the Development Studies Association* 20, no. 8 (2008): 996–1027.
- Baumard, Nicolas, Alexandre Hyafil, Ian Morris, and Pascal Boyer. "Increased Affluence Explains the Emergence of Ascetic Wisdoms and Moralizing Religions." *Current Biology* 25, no. 1 (2015): 10–15.
- Biraben, Jean-Noël. "Essai sur l'Évolution du Nombre des Mômes." *Population (French ed.)* (1979): 13–25.
- Caldwell, John C., and Thomas Schindlmayr. "Historical Population Estimates: Unraveling the Consensus." *Population and Development Review* 28, no. 2 (2002): 183–204.
- Carlos, Ann M., Donna Feir, and Angela Redish. "Indigenous Nations and the Development of the U.S. Economy: Land, Resources, and Dispossession." *Journal of Economic History* 82, no. 2 (2022): 516–55.
- Clark, Colin. *Population Growth and Land Use*. London: MacMillan, 1968.
- Clark, Gregory. "Review of Angus Maddison's *Contours of the World Economy*." *Journal of Economic History* 69, no. 4 (2009): 1156–61.
- D'Alpoim Guedes, Jade, Theodore C. Bestor, David Carrasco, Rowan Flad, et al. "Is Poverty in Our Genes? A Critique of Ashraf and Galor. The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *Current Anthropology* 54, no. 1 (2013): 71–79; *American Economic Review*, forthcoming.
- Deaton, Angus. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins, 1997.
- Denevan, William M. ed. *The Native Population of the Americas in 1492*. Madison: University of Wisconsin Press, 1992.

- Deng, Kent G. "Unveiling China's True Population Statistics for the Pre-Modern Era with Official Census Data." *Population Review* 43, no. 2 (2004): 32–69.
- Durand, John D. *Historical Estimates of World Population: An Evaluation*. Philadelphia: University of Pennsylvania, 1974.
- Federico, Giovanni, and Antonio Tena-Junguito. "How Many People on Earth? World Population 1800–1938." Maddison-Project Working Paper WP-16, November 2022.
- Frankema, Ewout, and Morten Jerven. "Writing History Backwards or Sideways: Towards a Consensus on African Population, 1850–2010." *Economic History Review* 67, no. 4 (2014): 907–31.
- Gennaioli, Nicola, and Hans-Joachim Voth. "State Capacity and Military Conflict." *Review of Economic Studies* 82, no. 4 (2015): 1409–48.
- Gervais, Raymond R., and Issiaka Mandé. "Comment Compter les Sujets de l'Empire?" *Vingtième siècle. Revue d'histoire* 95, no. 3 (2007): 63–74.
- Goldewijk, Kees Klein, Arthur Beusen, Jonathan Doelman, and Elke Stehfest. "Anthropogenic Land Use Estimates for the Holocene–HYDE 3.2." *Earth System Science Data* 9, no. 2 (2017): 927–53.
- Greene, William H. *Econometric Analysis*, 8th ed.. London: Pearson, 2018.
- Guinnane, Timothy W. "We Do Not Know the Population of Every Country in the World for the Past Two Thousand Years." CESifo Working Paper No. 9242, Munich, Germany, 2021.
- . "Replication package." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2023-03-26. <https://doi.org/10.3886/E187426V1>.
- Hyslop, Dean R., and Guido W. Imbens. "Bias from Classical and Other Forms of Measurement Error." *Journal of Business & Economic Statistics* 19, no. 4 (2001): 475–81.
- Iyigun, Murat. "Luther and Suleyman." *Quarterly Journal of Economics* 123, no. 4 (2008): 1465–94.
- Maddison, Angus. *The World Economy: A Millennial Perspective*. Paris, France: OECD, 2001.
- . *Contours of the World Economy, 1–2030 AD: Essays in Macro-Economic History*. Oxford: Oxford University Press, 2007.
- McEvedy, Colin, and Richard Jones. *Atlas of World Population History*. Harmondsworth, Middlesex, England: Penguin Books Ltd, 1978.
- Nunn, Nathan. "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics* 123, no. 1 (2008): 139–76.
- Nunn, Nathan, and Nancy Qian. "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." *Quarterly Journal of Economics* 126, no. 2 (2011): 593–650.
- Palma, Nuno, Jaime Reis, and Mengtian Zhang. "Reconstruction of Regional and National Population Using Intermittent Census-Type Data: The Case of Portugal, 1527–1864." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53, no. 1 (2020): 11–27.
- Putterman, Louis, and David N. Weil. "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics* 125, no. 4 (2010): 1627–82.
- Rogers, Alan R. "Evolution of Time Preference by Natural Selection." *American Economic Review* 84, no. 3 (1994): 460–81.

- Shiue, Carol H., and Wolfgang Keller. "Markets in China and Europe on the Eve of the Industrial Revolution." *American Economic Review* 97, no. 4 (2007): 1189–216.
- Voigtländer, Nico, and Hans-Joachim Voth. "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe." *Review of Economic Studies* 80, no. 2 (2013): 774–811.
- Vos, Jelmer. "Work in Times of Slavery, Colonialism, and Civil War: Labor Relations in Angola from 1800 to 2000." *History in Africa* 41 (2014): 363–85.
- Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press, 2010.
- Wrigley, E. A., and Roger Schofield. *The Population History of England 1541–1871*. Cambridge, MA: Harvard University Press, 1981.

Datasets Used in This Paper

- U.S. States: Haines, Michael R., and Inter-university Consortium for Political and Social Research. Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. <https://doi.org/10.3886/ICPSR02896.v3>
- MJ database. Data reported in MJ as provided by James Fenske.
- Nunn (2008) replication: <https://scholar.harvard.edu/nunn/pages/data-0>, accessed 22 January 2022.
- Maddison data: <https://www.rug.nl/ggdc/historicaldevelopment/maddison/original-maddison>, accessed 22 January 2022
- WB data: <https://data.worldbank.org/indicator/SP.POP.TOTL>, accessed 22 January 2022.
- Nunn and Qian (2011) replication: <https://scholar.harvard.edu/nunn/pages/data-0>, accessed 22 January 2022.
- Ashraf and Galor (2011) replication: <https://www.aeaweb.org/articles?id=10.1257/aer.101.5.2003>, accessed 22 January 2022.
- Ashraf and Galor (2013) replication: <https://www.aeaweb.org/articles?id=10.1257/aer.103.1.1>, accessed 22 January 2022.