How much more unequal? Consistent estimates of the distribution of wealth in the United States between 1774 and 1860

Martin Shanahan*, Margaret Corell

School of International Business, University of South Australia, North Tce., Adelaide, South Australia 5000, Australia

Received 1 June 1999

Abstract

Current estimates of long trends in the distribution of personal wealth in the United States combine a number of different studies. However, the trend estimates are open to challenge because of differences in methods of estimation between individual studies. In this article, a sample set from the 1860 census is analyzed and the distribution of wealth among different subsets of the population is described. Holding constant the method of estimation, we conclude that the apparent rise in inequality in the United States between 1774 and 1860, as measured using the Gini coefficient, is overstated by 0.1. © 2000 Elsevier Science Inc. All rights reserved.

JEL classification: N01; N31

Keywords: Wealth distribution; Estimation; Census data; Probate data

1. Introduction

Analysis of the distribution of wealth over time in the United States faces a number of complications. Records are diverse and inconsistent. They are frequently incomplete. Those data sets that exist for the period before the 1850s typically cover unrepresentative samples of the population. Further, the macroeconomic and demographic factors that affect the distrib-
tion of wealth change considerably in importance, as would be expected over 300 years. Patterns of settlement and migration, factor markets, price levels, emancipation, wars, and other social upheavals also impose their unique effects on any possible underlying wealth levels and trends.

Much work has been done to gather and analyze the data on personal wealth. There exist dozens of individual studies of wealth in particular communities or cities at particular moments of time. These studies have typically used extant records and the appropriate adjustment techniques to produce a singular wealth estimate for that location. Several national “snapshots” of the distribution of wealth also exist, the majority beginning with the censuses in the mid-nineteenth century (Soltow, 1975). The earliest complete study is by Jones (1977, 1980), who compiled a data set from a range of locations to construct a “national” estimate for the American Colonies in 1774. Williamson and Lindert (1980a,b) and Lindert (1991, 1998a,b) finally brought together the disparate threads of this research to provide long-run estimates of wealth trends. Their views on these trends have served as the foundation for current economic interpretations.

In an earlier article (Shanahan & Corell, 1998), we argued that differences in estimation methods mean existing estimates of long trends in the distribution of personal wealth in the United States are open to challenge. Specifically, differences between databases, adjustment methods, and geographic coverage suggest that pre-1850 estimates should not be directly compared with post-1850 wealth studies.

This article attempts to quantify the importance of this argument by examining current estimates for 1774 and 1860. These dates are of interest for a number of reasons, not the least of which is the ongoing debate about the trend in the distribution of wealth over this period (Soltow, 1971, 1984, 1989; Shammas, 1993; Lindert, 1998a,b). Second, current trend estimates rely heavily on the wealth studies for these two dates. Third, samples from the 1860 census have recently been made available electronically, greatly increasing the access of researchers.

In this article, we first illustrate the divergence in views on wealth distribution over the relevant period and then, using the 1860 census, illustrate the impact different wealth-holding units have on distribution estimates. We then describe an experiment by which we produce wealth distribution estimates for 1860 that are technically consistent with those produced for 1774. Finally, we examine the implications of our results and suggest future research.

2. A divergence of views

The 1774 study by Jones (1970, 1972, 1978, 1980) serves as the benchmark for wealth distribution estimates in America at the time of the revolution. Her study combined carefully sampled and weighted probate records with wealth estimates for the non-probated to produce a detailed picture of the overall distribution of wealth. She found personal wealth distributed relatively equally.

Soltow’s (1975) study estimates the distribution of wealth among adult males between 1850 and 1870. He meticulously samples census files to produce an estimate of the distribution of personal and real estate among a randomly selected set of adult males (aged
20 or over) in 1860 and 1870. He concludes that the distribution of wealth is highly unequal. The findings of these two studies are summarized in Table 1.

Consistent with Table 1, the current widely held view on trends in the distribution of wealth from the late eighteenth to the early twentieth centuries is summarized in Williamson and Lindert (1980b, pp. 62–63). Their work, which carefully and thoughtfully pieced together dozens of separate studies, was clear in its conclusions.

Between 1774 and the outbreak of the Civil War, the distribution of wealth appears to have undergone episodic change. Our nationwide estimates point to a near tripling in the ratios of the average wealth of the top 1 percent and 10 percent of wealth holders to the average wealth of all other groups ... regional estimates suggest most of the ante-bellum shift to wealth concentration occurred from the 1820s to the late 1840s ... We still know little about wealth inequality trends within the long period from the Civil War to World War I ... For the half-century after 1870 we are in the dark so we cannot with confidence identify peak inequality with 1929, 1914, or 1860.

Soltow, however, disagrees. Following his study of census records in the mid-nineteenth century, he examined samples of census records of real estate for 1798. He concludes, ... my investigations of the inequality of wealth in the United States show strong inequality in 1798, with about half of adult males having wealth, and with the Gini coefficient for wealth holders being about 0.6. Little change in economic inequality took place from 1771 to 1798 or between 1798 and 1860. Possibly, there was a slight increase in inequality in the latter period, since the proportion of people owning property did decrease a little during the two generations between 1800 and 1860. Nevertheless, my working hypothesis is ... that inequality of wealth in America remained fairly constant for the century preceding the Civil War (Soltow, 1989, p. 243).

This view of wealth trends, however, remains in the minority. For example, Lindert (1998a) considers the hypothesis of long-run inequality runs counter to contemporary

<table>
<thead>
<tr>
<th>Top (%)</th>
<th>1774 Total estate</th>
<th>1860 Total estate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Free adult males</td>
<td>Free households</td>
</tr>
<tr>
<td>1</td>
<td>13.2</td>
<td>12.6</td>
</tr>
<tr>
<td>10</td>
<td>54.3</td>
<td>49.6</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>94</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>99</td>
</tr>
<tr>
<td>Gini</td>
<td>0.632</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Note: The figures for 1774 are based on Lindert’s recalculation of Jones (1977). Sources: Adapted from Soltow (1975, Tables 4.2 and 4.3, pp. 99, 103) and Lindert (1998b, Table 3).
commentators’ perceptions of inequality in America and Britain. The hypothesis is also difficult to sustain in the light of other indirect evidence on inequality, such as the relative concentration of land ownership or a comparison of farmland prices and wage ratios. Finally, it is difficult to compare directly Soltow’s estimates for 1850 to 1870 with his less complete estimates of 1798.

Clearly, the colonial wealth study by Jones and census-based studies of 1860 and 1870 by Soltow are pivotal to the debate on wealth trends between 1774 and 1860. Both produce comprehensive estimates of the distribution of wealth. It is our conjecture, however, that differences in databases and the associated methods of wealth estimation can explain much of the measured increase in inequality between 1774 and 1860.

3. The impact of changing measurement units when estimating wealth from the 1860 census

The importance of the measurement unit in affecting distribution estimates has long been acknowledged in inequality studies (Cowell, 1977; Atack & Bateman, 1981a,b; Atkinson, 1983, pp. 33–60; Shammas, 1993). This issue is particularly important when examining distribution estimates over time.

Recently, historians at the University of Minnesota have produced electronically accessible census data for many censuses, including 1860 (Ruggles & Sobek, 1997). The accessibility of these data (the Integrated Public Use Microdata Series, or IPUMS) and the care with which they were sampled from the original census facilitates an examination of issues such as the impact of wealth units on the overall distribution estimates. Using a sample from the 1860 census, it is possible to compare the impact on distribution estimates and measures of inequality such as the Gini coefficient.

Table 2
Number of cases in sample from 1860 census

<table>
<thead>
<tr>
<th>Characteristics of cases in sample</th>
<th>All records</th>
<th>Adults</th>
<th>White adults</th>
<th>Households</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>11,495</td>
<td>5,584</td>
<td>5,401</td>
<td>2,271</td>
</tr>
<tr>
<td>Male</td>
<td>5,860</td>
<td>2,903</td>
<td>2,816</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5,635</td>
<td>2,681</td>
<td>2,585</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>51.42%</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>20–44</td>
<td>35.18%</td>
<td>72.42%</td>
<td>72.41%</td>
<td></td>
</tr>
<tr>
<td>45 +</td>
<td>13.40%</td>
<td>27.58%</td>
<td>27.59%</td>
<td></td>
</tr>
<tr>
<td>Mean total wealth</td>
<td>$695.16</td>
<td>$1,418.28</td>
<td>$1,457.57</td>
<td>$3,518.65</td>
</tr>
<tr>
<td>Median total wealth</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$656.00</td>
</tr>
</tbody>
</table>

Note: Adult ≥ 20 years of age.
A random sample of 11,495 observations was selected from the IPUMS records of the 1860 census. Table 2 provides a summary of the cases in the sample. The sample reflects a population that mostly consists of free whites, with comparably high levels of average wealth and a median wealth of zero. The age, race, and sex characteristics of this sample are not inconsistent with those of Soltow’s (1975, pp. 99, 103) sample.

With this data set, it is possible to examine the impact a difference in measurement unit may have on the estimated distribution of wealth. Table 3 shows the results. For example, the distribution of wealth among all adults is measured by a Gini coefficient of 0.91, while reducing the measurement unit to adult males only reduced the coefficient to 0.85. White adult males (who make up the majority of the individuals recorded with wealth) also recorded a Gini coefficient of 0.85 while measuring the distribution of wealth by household unit reduced the coefficient further to 0.81.

Table 4 reports estimates of the distribution of wealth in America in 1860 calculated by three different researchers using different sample sets and measurement units. Comparisons of these results with those on the previous table reveal that our distribution estimates are broadly consistent with these other studies. Also in line with the calculation in Table 3, the measured level of inequality among adult males is greater than among households.

Exercises such as this are useful in drawing attention to the need for consistency of measurement units when comparing wealth distributions over time. It is also important to adopt consistent databases and estimation techniques.

---

1 The IPUMS records for 1860 are a one in 100 random sample of households drawn from the 2 million census pages. The IPUMS records for 1860 weight black people at twice the standard rate. The data are stratified by geographic location to produce a more even geographic distribution of cases than expected from a true random sample (Ruggles and Sobek, Vol. 1, Chap. 3 p. 3).
4. An experiment

If differences in technique affect the estimated distribution of wealth, then it may be possible to identify the magnitude of this impact by using a consistent method of estimation. The lack of a census for 1774 prohibits the adoption of a census-based method for that year, but it is possible to adjust the census data for 1860 to produce a “probate-based” estimate. However, producing a set of data that mimics probate records requires that data from the census be ignored. In effect, the experiment involves “turning good data into bad” (Romer, 1986, p. 3) by discarding information about the overall distribution of wealth and constructing an estimate from a much smaller base.

Probate-based estimates of the distribution of wealth are predicated on a number of assumptions. It is assumed that the amount of wealth recorded against the name of a person who died is essentially a “snapshot” of their wealth, the moment selected by the randomness of death. As those who are probated are typically more likely to be older, white males, these factors are “adjusted” to produce a group whose age and sex structure matches the living population. In essence, the adjustment produces a group of people whose characteristics match those who will be probated when they die—a select sample of the population. The other group in the population consists of those people who will not be probated when they die. These people, too, held some wealth and in order to produce a wealth estimate that covers everyone, some distribution of wealth must be attributed to this group. The final wealth

Table 4
Estimates of the distribution of wealth in America in 1860

<table>
<thead>
<tr>
<th>Percentage of wealth held by top</th>
<th>Soltow (1975)</th>
<th>Attack and Bateman (1981a,b)</th>
<th>Steckel (1990)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>US</td>
<td>West</td>
<td>East</td>
</tr>
<tr>
<td>1%</td>
<td>29.0</td>
<td>11.2</td>
<td>12.39</td>
</tr>
<tr>
<td>5%</td>
<td>57.0</td>
<td>30.6</td>
<td>31.72</td>
</tr>
<tr>
<td>10%</td>
<td>73.0</td>
<td>46.0</td>
<td>66.8</td>
</tr>
<tr>
<td>20%</td>
<td>86.0</td>
<td>64.0</td>
<td>65.7</td>
</tr>
<tr>
<td>Gini</td>
<td>0.832</td>
<td>0.62</td>
<td>0.65</td>
</tr>
<tr>
<td>Unit of measurement</td>
<td>Adult males</td>
<td>Households</td>
<td>Male household heads with children aged &gt;10</td>
</tr>
<tr>
<td></td>
<td>≥ 20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Soltow (1975, Tables 4.2 and 4.3, pp. 99; 103); Attack and Bateman (1981b, Table 1, p. 125); Steckel (1990, Table 1, p. 277).

2 An alternative to modifying the census data for 1860 is to estimate the distribution of wealth from a set of probate records collected for the same year. One study, which directly compares census-based wealth distribution estimates with probate-based estimates, is by Shanahan (1995) where the Australian wealth census estimates of 1915 are compared with probate-based estimates in one state for the same period.
estimate is therefore a combination of the distribution of wealth among a “known” group and an unknown group.

Consequently, our method involves bias sampling the census data of 1860 to produce a “probate-based sample.” We then construct a wealth distribution estimate, which “matches” the approach of Alice Hanson Jones for 1774. Effectively, we work “backwards” from the census towards a probate sample from which to construct wealth estimates. The experiment removes the influence of inconsistent estimation techniques to produce a more reliable description in the trend of inequality over time.

If we assume the census provides information about the entire living population, we can sample this in such a way as to select a group of people who will be probated when they die. First, a perfectly random sample is drawn from the census; in this study, a sample is taken from the IPUMS records. The mean characteristics and attributes of this sample should, within the bounds of our confidence limits, match that of the census. In theory, this “sample” is equivalent to the ideal “population” at the end-point of an estimation process that began from probate records.

Next, we bias sample the set, with a view to selecting those people who would have been probated when they died. This set is a replica of the product that results when probated estates are “multiplied-up” to resemble more closely the sex and age distribution of the living. 3

A person was more likely to be probated if he was older, white, and wealthy (Main, 1974; Smith, 1975). Following Jones (1980, p. 39), only 55.5 percent of potential wealth-holders were likely to be probated when they died. The initial census sample was reduced to include only white, free adults (aged 20 or over) as the potential wealth-holders. To produce a sub-sample that represents those people who would have been probated requires a decision-making rule to produce a biased sample of the initial data set. The simplest is to rank by wealth and select the top 55.5 percent as the set of people most likely to be probated at death. 4

The experiment, however, is not complete. Assuming we know less about the population than we really do, we must estimate the wealth of that portion of the population who would not have been probated when they died. Following the method adopted by

---

3 In theory, we could take this exercise further, and divide this set by the appropriate mortality rates attributable to each sex and age group to produce a very small “probate” sample. We did not continue to this end-point. Such a probate set would, however, be equivalent to the unadjusted data sets used in several pre-colonial wealth studies to produce wealth estimates; for example, Warden (1976).

4 As an alternative to simply selecting the top wealth-holders, we also examined the proportion of the probated group in 1774 who left an amount of wealth higher than the median of the whole population (the median of the \( w^*B \) estimate in Jones, 1978, p. 2105). This figure amounted to 66.4 percent. We thus randomly sampled from the set of white adults aged 20 years and over, to produce a set containing 55.5 percent of all white adults, and where 66.4 percent of these had wealth greater than the median, and 33.6 percent less than the median. The aim was to produce a probate sub-sample that was more thoroughly mixed that occurs with a simple ranking procedure, but that still represented a data set biased toward those who would be probated when they died.
Jones (1978), we attribute one-quarter of the average wealth of the “probated group” to
the non-probated.\textsuperscript{5}

The distribution of this wealth among the non-probated can take several forms. The
simplest is to allocate the one-quarter average wealth quantity only to the 44.5 percent of
the population previously excluded. This assumes that most, if not all, free white adults left
some wealth, although they were not probated. A second method involves distributing the
one-quarter wealth average to both probated and non-probated, on the assumption that
some people who left wealth were missed by the probate process, while there were still
people (very much fewer than assumed above) who left virtually no wealth.\textsuperscript{6}

The final wealth estimates, which combine the distribution of wealth among the “known”
multiplied-up probate group and the unknown group, are the result. These estimates for 1860
are directly comparable to those produced by Jones for 1774.

Based on the probate method of calculating the distribution of wealth, the preferred
estimate for 1860 reveals the top 1 percent of wealth-holders possessed 35.2 percent of the
nation’s personal wealth. The top 10 percent were estimated to hold 70.81 percent and the top
20 percent, 80.22 percent. The overall Gini coefficient was 0.723 for this distribution.

Comparing these results with those in Table 3, it is clear that the estimation of the distribution
of wealth using the adjusted probate method lowers the estimated level of inequality and the
recorded Gini coefficient. For example, the distribution of wealth among all free white adults
taken from the census sample is summarized by a Gini coefficient of 0.912. The preferred Gini
coefficient for white adults using the probate method of estimation is 0.189 less, at 0.723.

By holding the estimation technique constant, it is also possible to examine where changes
in the distribution of wealth may have occurred over time. Our inferences, however, can only
ever be tentative, given their foundation on a single sample of the 1860 census and the
artificial manner in which the distributions have been derived. Nevertheless, it would appear
that the primary shift that occurred between 1774 and 1860 occurred in the upper end of the
wealth distribution. This conclusion is consistent with the view of Williamson and Lindert
(1980a,b) that wealth concentrations increased among the wealthy. Although this experiment
concurs with their argument, the increase in inequality is not as large as their work suggests.

The results of this experiment suggest measured inequality in the distribution of wealth
increased between 1774 and 1860. The amount of increase, however, would not appear to be
as great as previously thought. The comparison should not be between distributions that
report an increase in the Gini coefficient among adults from 0.632 in 1774 (based on fully
adjusted probate records) to 0.832 in 1860 (based on a sample of adult males in the census).

\textsuperscript{5} Jones (1978) uses a complicated “weighting” procedure where she allocates an average wealth equal to 1/2
that of the probated estates in one region, while in others she allocates the non-probated 1/4 the average wealth
of the probated. She states (p. 1867): “In general [the allocation] could be varied to any ratio which upon further
research into the relation between total wealth-holder deaths and total numbers of cases of probated estates, the
investigator would find suitable.” In this study, we allocate 1/4 the average wealth of the probated set to the
non-probated.

\textsuperscript{6} The second method involves allocating the aggregate wealth distributed above across all those with lower
wealth-holdings (probated and non-probated) and results in an average wealth level among the recipients equal
to less than one-quarter the wealth of the probated group.
Rather, the comparison should be between two distributions of potential wealth-holders, based on probate techniques, which reveal an increase from 0.632 in 1774 to 0.723 in 1860.

The experiment suggests that possibly up to 0.1 of the change in the Gini coefficients might be attributable to differences in estimation techniques. That is, either the probate-based approach as adopted and applied by Jones understates inequality, or the census sample of Soltow overstates inequality. We suspect the former, because of the need by Jones to use a crude estimate for wealth-holdings when estimating the distribution of wealth in the lower tail.7 The

---

"Note: Age adjusted"—probate records adjusted for age and sex to represent those who would be probated when they died. It does not include the non-probated.

"Includes non-probated" is a complete distribution including adjustments for the non-probated. For the 1774 figures, this corresponds to Jones’ \( w^A \) weighted distribution: for 1860 it assumes “non-probated” held 1/4 average wealth of probated set.

"Preferred estimate" 1774 figures correspond with Jones’ \( w^B \) weighted distribution, while the figures for 1860 assume non-probated and lower wealth-holders of probated records were omitted. The aggregate wealth redistributed in column 5 was redistributed across all lower wealth-holders.

"Third estimate" is calculated by extracting a random sample of all adults equal to 55.5 percent of the total white adult population. The parameters of this subset are that 66.4 percent hold wealth greater than the median for the entire population. The 1/4 of the average wealth of this sub-group is then redistributed to all lower wealth-holders. The figures for 1860 are based on white adults (≥20 years).

Source: Jones (1978, Vol. 3, Table 7.176, p. 2112) and authors’ calculations.

---

7 This should not be taken as a criticism of Jones’ work. Not only does she make her assumptions very clear, and remark on the need for better figures, but also, she had no other information (such as a census) by which to calculate estimates of the missing wealth. Only with more information could Jones’ estimates have been made more precise.
importance of how wealth is distributed in the tail of the overall distribution is highlighted in Table 5 in the last column. Here, one-quarter of the average wealth of the probate subset is distributed to all wealth-holders recorded as possessing less than this one-quarter average. The impact on the lower tail of the distribution and the overall inequality measure is dramatic, lowering the Gini coefficient to 0.644, or virtually the same coefficient as in 1774!

The current view of a steep upward trend in inequality across the nineteenth century is at least, in part, the consequence of beginning from a comparatively “low,” probate-founded database, and comparing this with a comparatively “high” census database.

5. Conclusions

Distribution estimates are sensitive to a wide range of factors. Alternative databases and their associated distribution estimation techniques have been shown to be an important variable when examining changes to the distribution of wealth over time.

Ideally, comparisons of the distribution of wealth over time should only be made between like data sets and similar units of measurement. Probate records are available across several centuries, with different levels of representativeness and complementary information. Future research should re-examine existing probate-based estimates with a view to using consistent adjustment methods, particularly when estimating the wealth of the non-probated. Where this is not possible, some calibration should be employed to adjust for the impact that usage of different databases may have upon wealth estimates.

In the case of wealth trend estimates for the period 1774 to 1860, we find measures of inequality for 1860 to be higher in part because calculations were undertaken from a census sample. The increase in inequality resulting from a comparison of a probate-based estimate against a census-based is approximately 0.1 as measured by a Gini coefficient. This suggests that while inequality did increase over the century before 1860, the difference in inequality may not be as large as current interpretations by Williamson and Lindert (1980a,b) suggest. Indeed, the current debate as to whether inequality remained high across the period, or whether it increased from a more equal distribution in 1774, may as much be the result of differences in databases, estimation techniques, and units of measurement, as to real differences in inequality.

Acknowledgments

The authors would like to thank William Becker, Peter Lindert, Richard Steckel, Mervyn Lewis, and Ian McLean for assistance at different stages of this article, together with suggestions by participants at the Third World Congress of Cliometrics, and the XIth International Congress of the International Economic History Association. Naturally, we are responsible for all errors and omissions.
References


