This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: American Economic Growth and Standards of Living before the Civil War

Volume Author/Editor: Robert E. Gallman and John Joseph Wallis, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-27945-6

Volume URL: http://www.nber.org/books/gall92-1

Conference Date: July 20-22, 1990

Publication Date: January 1992

Chapter Title: Inequalities in the Standard of Living in the United States,1798-187

Chapter Author: Lee Soltow

Chapter URL: http://www.nber.org/chapters/c8009

Chapter pages in book: (p. 121 - 172)

3 Inequalities in the Standard of Living in the United States, 1798–1875

Lee Soltow

A presentation of the standard of living for the United States in 1800 or 1860—or even for all intervening years—by using an average value per person of any economic variable yields, at best, a partial quantification of the matter. Ideally, one would also like to study a time series of averages just for the rich and just for the poor. Even better would be tables stating shares for the five quintile ranges, similar to those that have been available for income since 1947.

How wonderful it would be if there were complete distributions of wealth and income for each of the sixty years prior to 1860. One then could make statements about how the poor fared, relative to the rich, with the onset of the industrial revolution. In what sense might the rich have grown richer? Perhaps it was middle groups that gained relative to those above and below them when changes in the industrial and occupational structures occurred, as depicted by other participants in this conference on the standard of living.

This paper is a statement of some scattered and irregular distributions for years for which data are available. I feature tables for wealth and income, and distributions for saving (wealth), house values and rent, food consumption and nutrition (as reflected in heights of males and farm production), and clothing expenditure (as shown by home-woven yardage). I place special emphasis on shares of the poor, on households and individuals below the fortieth percentile for any particular variate. Changes in relative inequality will be emphasized, although this is possible in only three situations.

3.1 Sources and Findings

Very few distributions exist for the United States in the nineteenth century. Most prominent among them are those for wealth in 1798, 1850, 1860, and

Lee Soltow is professor of economics at Ohio University.

1870, as revealed in the censuses of wealth for those years. A strategic frequency classification of real estate values in Ohio in 1835 can serve as an intervening quantification in sketching changes in the hierarchy of values. These five distributions appear presently in chart form. Further perspective will be provided with a chart showing wealth distributions for those of older age in 1850; these are individuals whose fortunes were formed between 1800 and 1850. The central features of the two charts are very similar, and one must generally conclude that wealth inequality among free persons with real and personal estate changed very little in the first three-quarters of the nineteenth century.

The wealth data for the nineteenth century suffer one glaring defect; they tell us nothing about the 40 to 50 percent of adult free males and the 30 to 40 percent of families who possessed no wealth other than clothing, tools, a little furniture, and perhaps some farm animals. It is of fundamental importance to know whether this group expanded or contracted before the Civil War, and whether this group's condition either improved or deteriorated. Unfortunately, there are no time series or any selected points in the period on which one can focus in depicting change for lower groups. In analyzing conditions we must begin somewhere; to this end I will present in various sections of this paper some data concerning the lower 40 percent of people.[1] The most poignant statistics in this respect will be those dealing with persons living in shanties in Ontario in 1851. Some inkling of conditions on marginal farms will be presented for South Carolina in 1850. The extent of deprivation, as it appears in data for education and family formation, will be given for the United States in 1870; these frequency tables do suggest deprivation. Persons without wealth reported both significantly lower marriage rates and fewer children.

The data for wealth suggest that half of adult free males held real estate in 1798, and that this proportion decreased to .41–.43 in the period from 1850 to 1860. Yet one must not make too much of this matter, since surely the long-run trend in the occupational shift from agriculture to manufacturing and services was operative during the century. What we really need is a number of *income* distributions for the total labor force similar to those that are available for years after 1947. To do this demands using statistics, or possibly proxies, for *all* individuals in the country. Two main sources of information comprising certain aspects of such a comprehensive coverage are housing values or rents and heights of army recruits.

I have made an estimate of the distribution of income among families and individuals for 1798 using the splendid information available from the dwelling tax of that year coupled with information on the number of persons in dwellings. This distribution, as pictured in figure 3.1, suggests an important ordering principle: the inequality in wealth or real estate among wealthholders (with a Gini coefficient, *G*, of about .6) approximates the degree of inequality

1. Amartya Sen (1976) stresses the importance of the degree of inequality in the lower tail.

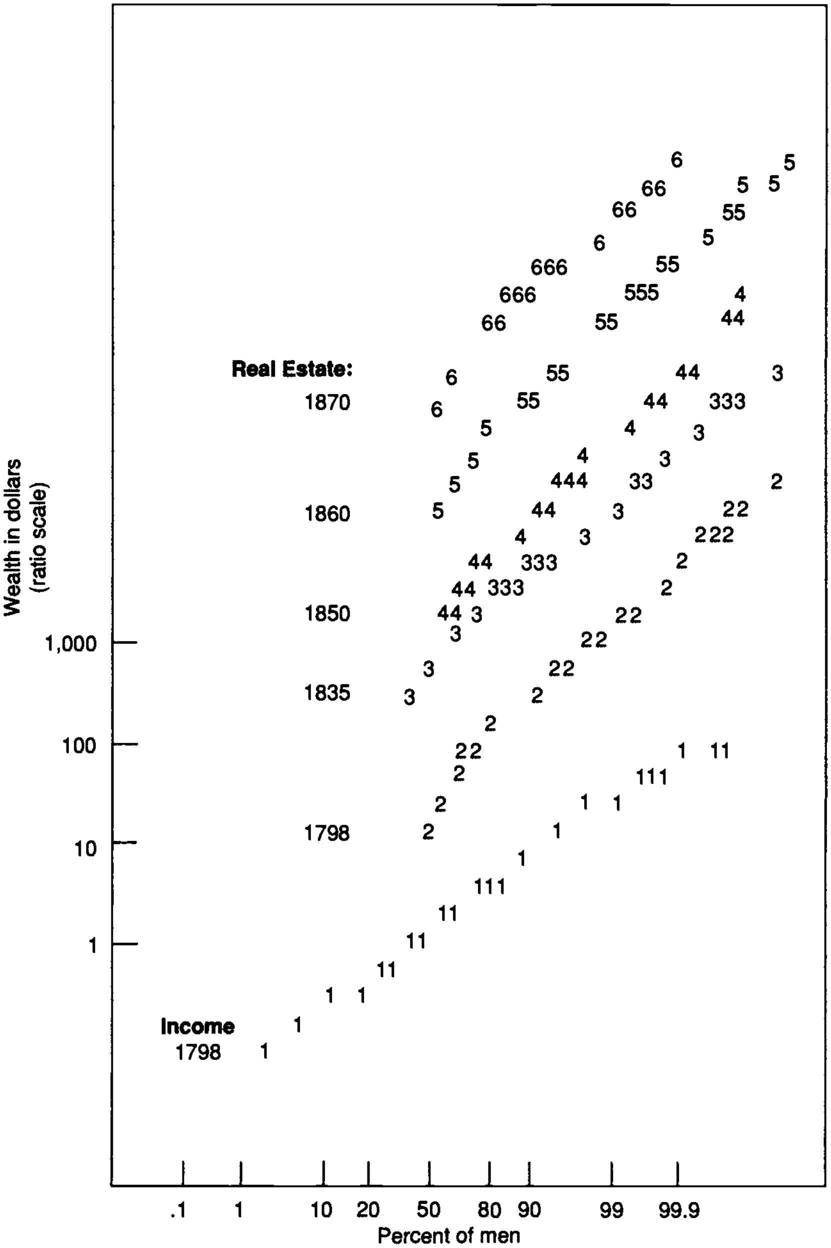

Fig. 3.1 A tier chart of the distribution of real estate among adult free males in the United States in 1798, 1850, 1860, and 1870 (whites), and in Ohio in 1835 (a lognormal probability chart)
Source: See table 3.1.

found in housing and dwelling-derived income among all families and unrelated individuals (G = .6), at least in 1798.

The 1798 wealth-dwelling pattern is suggestive of a similar pattern for 1850 and 1860, one with little change in the shapes of relative distributions in the half-century. Nevertheless, there are no dwelling-value distributions available for the country that can serve as checks until 1980, a date generally beyond the scope of this paper. The 1798 and 1980 distributions strongly point to decreasing inequality of income, but one suspects that the drop occurred largely in the twentieth century.

In New York State there were censuses of housing in 1865 and 1875 as well as in 1798; of necessity, one must turn to these data in order to make statements about changes in income distribution. A small sample from the manuscripts in 1865 indicates very little change—perhaps some lessening in inequality. Fortunately, the distribution of housing values in 1875 has been tabulated. A major section of this paper is devoted to comparing the housing distributions of New York in 1798, 1875, and 1980. The frequency tables clearly show little change before 1875; the housing share of the lowest groups remained relatively constant. Only after that date does this share rise dramatically. I present some further dimensions of housing in later sections, stressing the variability in lower income shares in various regions of the country. Housing data for Boston, Amsterdam, and Antwerp point, in a most rudimentary fashion, to similar shifts in relative inequality in both Boston and Amsterdam.

Heights of all army recruits are available for the periods centering on 1818, 1864, and 1918. One can argue that food consumption and general nutrition affect not only mean height, but also the median, first quartile, first decile, and so forth, values of height distributions at the different dates. I investigate relative distributions and offer Lorenz curves of height for the above dates. Very little change, if any, is found, and perhaps very little change after 1918 as well. Shares of lower groups remained constant. The 1812–63 food-height experience is consistent with that for housing in the nineteenth century but not in the twentieth century. Some might assert that U.S. standards of food consumption always were sufficiently high that one might not be able to discern the effects of differences in consumption by studying relative height. Others might assert that relative heights essentially do not reflect income distribution. Yet differences in relative heights decreased in Scandinavia, Holland, and Amsterdam during the last century.

Historical information is available for distributions of housing, perhaps for food, and for saving (as reflected in wealth). Not so certain are aspects of clothing distribution. One suspects, however, that the shift from household to factory production of cloth (and of clothing) benefited lower-income groups relative to upper groups. Some hints of this process are indicated in data from the New York State censuses of 1825, 1835, and 1845, at least from the standpoint of home production of cloth yardages; I will present a distribution from sample data of household yardage production. In my general summary I consider the consistency in inequality of saving as well as inequalities in expenditures on housing, food, and clothing combined as an income distribution.

What do the distributions for the nineteenth century tell us? One must conclude that relative inequality changed very little before the Civil War. Still to be revealed is the fact that lower-income groups had only a small share of resources in any one year, at least in many respects.

3.2 Wealth Distribution

For the most part, individuals probably had a better understanding of their assets than of their incomes, particularly in a rural setting. In any case, strong emphasis must be placed on the measurements of wealth. Our censuses of real estate, coupled with the censuses of population for various years in the nineteenth century, are truly unique.

3.2.1 Real Estate

The distribution of wealth in real estate among adult free males in the United States in 1798, 1850, 1860, and 1870 appears in figure 3.1, a depiction using lognormal probability paper. An exact lognormal curve plots as a straight line on this convenient chart form; it is desirable since many wealth and income distributions are approximately normal in shape when using logarithms of the variate. The greater the slope of the line (b), the greater the Gini coefficient of inequality (G); there is an exact relationship between b and G for straight lines on this chart.[2]

Calculations of the inequality coefficients and slopes are stated in table 3.1. The G and b estimates for wealth in real estate in 1798 and 1850–70 differ very little, and one reaches the conclusion that there was little or no change in inequality. But what about the intervening half-century between those census dates? The data for Ohio in 1835 prove to be a strategic source in filling this gap. Ohio's wealth in real estate reflected settlement a generation after it achieved statehood. In 1800–1835, its pattern of inheritance reflected in part the inheritance patterns of settlers from Pennsylvania, Virginia, and other states. In his trip to Ohio in 1831 and 1832, Tocqueville (1966) singled out Ohio for its tremendous developmental efforts. In 1840, Ohio's population accounted for 17 percent of the country's population. From the standpoint of real estate ownership, Ohio's distribution was similar to that in the nation as a whole.

Further insight into wealth patterns in 1840, 1830, and earlier can be ob-

2. Figure 3.1 is a lognormal chart whose horizontal scale is the standard normal deviate, z, stated here in probability form such as Prob ($z < -1.28$) = .10. A straight line on this chart is a lognormal curve, and I used the form $LL = a + bz$, where LL is the value at the lower limit of the class in the case of a frequency table; b is an estimate of the standard deviation using logarithms of the variate.

Table 3.1 **Summary of the Degree of Relative Inequality of the Distributions of Weal in Real Estate among Free Adult Males in the United States in 1798, 1835, 1850, 1860, and 1870**

	Wealth > 0			Wealth ≥ 0			Straight-Line Model		
	G	Mean	Number (thousands)	*G*	Mean	Number (thousands)	*b*	R^2	*Poir*
Real estate									
1870 whites, U.S.	.624	4,150	3,700	.833	1,850	8,300	1.79	.981	24
1860 free, U.S.	.649	3,500	3,000	.845	1,540	6,800	1.77	.995	31
1850 free, U.S.	.643	2,470	2,000	.848	1,046	4,800	1.83	.996	29
1835 Ohio[b]	.637	530	139	.799	294	250	1.73	.993	32
1798 free, U.S.[c]	.632	1,430	433	.818	708	878	1.78	.982	30
Income, dwelling-derived									
1860 free, U.S.				?		?			
1798 free, U.S.[d]				.631	348	878	1.26	.996	26

Sources: 1850–70 distributions are computed from spin samples described in Soltow (1975, 96), samp with sizes of 10,393 in 1850, 13,698 in 1860, and 9,824 in 1870. The 1835 distribution for Ohio incluc 164,962 property values derived from the tax duplicates in1835; properties are collated by owner's nai within a county to yield 138,785 owners. The 1798 wealth distribution is derived from a disproportion sample of 46,046 evaluations as described in Soltow (1984; 1989a, chap. 2). The dwelling-deriv income distribution stems from a sample of 39,890 dwelling values and nineteenth-century summa tables of ten dwelling-tax classes, as described in Soltow 1987a; 1989a, chap. 3, 42, 263, 264). I a preparing a computer tape of these six distributions and nine others that I will submit to the Int University Consortium for Political and Social Research.

Notes: G = Gini coefficient of realtive inequality; b = the slope of the line on lognormal probabili paper, shown in figure 3.1. Adult males are twenty-one years and older.

[a]Distributions were partitioned into classes with lower limits generally having a first digit of 1, 2, 5, (in the lowest range) 1, 2, . . . , 9. These classes or points, when plotted on figure 3.1, usually display one or two points at the lowest level that clearly were below the linear shape of all other points. eliminated these one or two before fitting the linear model.

[b]Owners' names were collated within counties, yielding wealth for 138,785 owners. See Soltow (1987 138).

[c]An upward collation adjustment has been made using an elasticity coefficient of 1.11, as suggested fro the Kentucky experience. See Soltow (1984, 450).

[d]This is my preferred estimate using an elasticity of dwelling expenditure with respect to income of 1. See Soltow (1987a, 184; 1989a, 247, 273).

tained by classifying persons in the 1850 census by *age*. For those 70 and older, 60–69, 50–59, and 40–49, their wealth in the 1850 census to a great extent reflects their activities in previous decades. Plottings of the distributions for each group are given in figure 3.2. Certainly there are neither discontinuities nor serious alterations in the slopes of the lines that would signal discontinuities in relative inequality. The results for 1860 and 1870 are similar; data for total estate (real and personal wealth) appear in the last three columns of table 3.2, but without considering the small number of persons with wealth from $1 to $99, or those with no wealth.

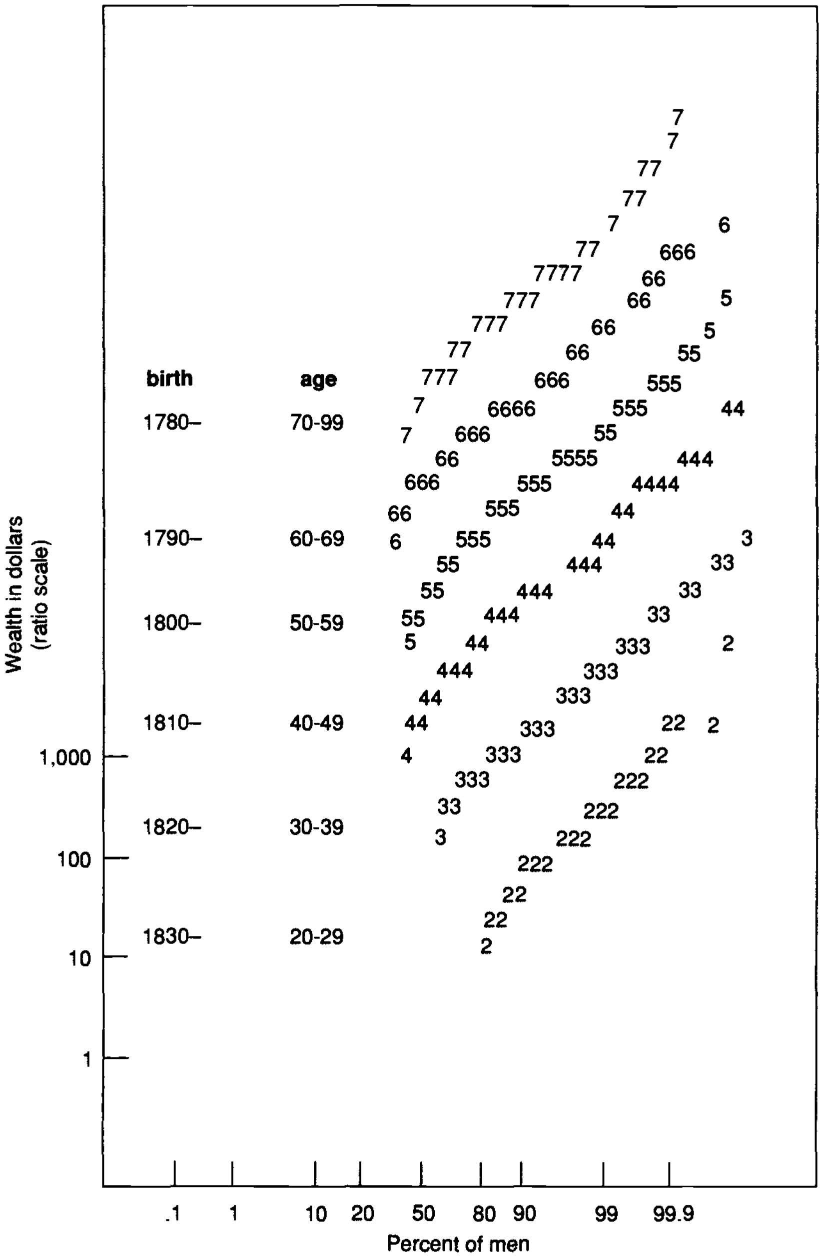

Fig. 3.2 A tier chart of wealth in real estate of adult free males in the United States in 1850, classified by age

Source: See table 3.1.

Table 3.2 **Wealth Distribution among White Males Twenty and Older in the United States in 1870, Classified by Five-Year Age Intervals**

Age	Number in Sample	Proportion with Wealth ≥ $100	Persons with Wealth ≥ $100				
			Number	Mean ($)	G	R^2	
20–24	1,756	.26	449	1,590	.66	.974	1.
25–29	1,381	.56	769	1,630	.63	.972	1.
30–34	1,147	.65	750	2,730	.68	.984	1.
35–39	1,105	.73	803	3,370	.66	.984	1.
40–44	858	.76	653	4,910	.66	.994	1.
45–49	788	.78	612	5,970	.66	.994	1.
50–54	705	.83	557	6,090	.71	.991	1.
55–59	415	.83	370	6,160	.65	.992	1.
60–64	373	.80	297	5,220	.64	.993	1.
65–74	426	.86	366	7,060	.72	.992	1.
20–29	3,137	.39	1,218	1,610	.64	.974	1.
30–44	3,110	.71	2,206	3,610	.67	.988	1.
45–64	2,311	.79	1,836	5,920	.67	.994	1.

Source: A sample of 10,235 males twenty and older, including 9,125 whites. The data are described Soltow and Stevens (1981, 221 n. 72).

Note: G = Gini coefficient of relative inequality; R^2 = coefficient of determination for the straight-li model with slope b fitted on lognormal paper.

3.2.2 The Poor in a Land of Wealth

The fundamental weakness in the wealth analysis is the absence of perspective concerning those without estates. In this respect, some headway may be made if we can determine the characteristics of those without wealth; such data are revealed particularly in the 1870 census. I choose to describe this year because of information on education, marriage, and number of children, derived from a study I made of literacy.

The proportion owning estates valued at above $100 is the crucial element in understanding the condition of persons who are, in some sense, in the lower half of the wealth distribution. In this connection, the proportions given for five-year age intervals in the second column of table 3.2 are strategic. The pattern that emerges is that the majority had wealth after age twenty-five, and that a strong majority enjoyed at least some wealth after age forty-five. The inequality of wealth only among those above $100 was similar at any age. It is as though those arriving at the bottom of the distribution in their older age appear in an orderly fashion, joining the lognormal distribution as it previously existed without being on some parallel line pitched at a lower level.

The aging process obviously leads to greater shares of wealth and probably of total income of persons below the median (see table 3.3). Part of this increasing equality arises from the enhanced probability of receiving inheri-

Table 3.3

Array of Wealthholders with Wealth ≥ $0	Share of Aggregate Wealth of All, by Age		
	20–29	30–34	45–65
Lowest 50% in array	.00	.0234	.0426
Lowest 40% in array	.00	.0074	.0275
Lowest 30% in array	.00	.0004	.0046
Lowest 20% in array	.00	.0000	.0000

tances as one becomes older (Soltow 1982). And part arises from the ability to share in capital gains once having acquired real estate.

The possibility arises that the above interpretation is misleading. Perhaps those without wealth die at older ages in far greater numbers than do those with wealth. Table 3.2 shows that the maximum number of holders appears in the age group 35–39, and that this number is not much larger than for those 25–29. Admittedly, many confounding factors arise when one compares class frequencies in this fashion; these include immigration and the effects of the Civil War.[3]

Of more direct importance would be the characteristics of poor and rich at any specific age. How much of a handicap arose from not having wealth or income? The census does reveal a few characteristics dealing with economic well-being. The first of these was the ability of the parent to send children to school, an expenditure in some ways as significant as that for food, clothing, and housing. The data of table 3.4 indicate that the children of the rich enjoyed attendance probabilities (SCH) one-third to two-thirds again as large as the probabilities for the poor. Though these are significant differences, to my mind they are relatively small. A system had arisen that gave benefits to the poor in significant numbers. This was not a society of aristocrats, with school attendance probabilities for the rich two or three times those of the poor.

Data from tables 3.2 and 3.4 can be combined for white males aged forty-five to forty-nine. About 22 percent of adult males had no wealth. The majority of their children aged ten to fourteen were in school. The record for those next in line above the poor was no better, while the children of the somewhat rich had school attendance probabilities only somewhat better than those for the children of the poor.

An alternative measure of the differences between poor and rich is derived from counts of marital status and the number of children born to a family.

3. Dr. William Ogle (1887, 650–51) compared death rates at various ages of the old English life table for the years 1838–54, with the Upper Class Experience Table. He found that the percentage surviving after thirty years was 60 in the former group and 70 in the latter for males thirty years of age. See also Humphreys (1887, 277) and Fogel (1986, 467).

Table 3.4 **Enrollment Rates in Schools (SCH) of White Children in the United States in 1870, Classified by Age of Child, Occupation of Father, and Wealth of Father**

Wealth Class of Father ($)	Age of Farm Children 5–9	Age of Farm Children 10–14	Age of Farm Children 15–19	Age of Nonfarm Children 5–9	Age of Nonfarm Children 10–14	Age of Nonfarm Children 15–19
All children						
0–99	.34	.56	.43	.44	.62	.24
100–999	.35	.56	.33	.52	.68	.32
1,000–9,999	.50	.73	.46	.56	.74	.44
10,000 and up	.62	.86	.67	.59	.83	.63
Children with fathers of age 45–49						
0–99	.42	.53	.44	.47	.67	.29
100–999	.39	.50	.24	.56	.64	.36
1,000–9,999	.53	.69	.52	.61	.72	.34
10,000 and up	.65	.91	.68	.68	.86	.58

Source: A sample of 19,117 children in 9,125 white families reporting fathers. The above data are for 12,312 children aged 5–19 and, of these, 1,867 with fathers 45–49. Also see Soltow and Stevens (1981, 221 n. 72).

Table 3.5 **Marriage Rates and the Average Number of Nuclear Children of White Males in 1870, Classified by Wealth**

Age	Proportion Married: Without Wealth	Proportion Married: With Wealth	Average Number of Children: Without Wealth	Average Number of Children: With Wealth
20–24	.21	.71	0.12	0.60
25–29	.43	.86	0.57	1.26
30–34	.59	.93	1.06	2.15
35–39	.64	.93	1.81	3.05
40–44	.65	.93	2.13	3.36
45–49	.77	.93	2.55	3.60
50–54	.64	.93	1.69	3.39
60–64	.71	.86	1.61	1.99

Source: A sample of 9,125 white males twenty and older and the 16,828 children of those having children enumerated as living with them. See also tables 3.2 and 3.4, particularly for the number of adult males, fathers and nonfathers, in each age interval; and Soltow (1982).

Surely the economic literature dealing with marriage rates from at least the time of Malthus has stressed that the poor should postpone marriage. Here was a most vital element of the economic well-being of the poor, as contrasted to that of the rich, and I must ask what the differential marriage rates for age-specific groups were, considering the data of table 3.5.

Marriage rates among those with wealth reached a maximum of 93 percent, with remarkable constancy from age thirty to age fifty-nine, as stated in the table. There was a very substantial difference in the marital status among the relative poor and rich, particularly for those in their twenties. The poor obviously were forced to postpone marriage, much as Malthus might have wished. Yet they were also penalized in their fifties when about one-third remained single, as determined by subtracting their marriage rates from the 93 percent upper asymptote of those with wealth. One can argue, of course, that there is an element of tautology in the data. An individual with wealth owned a farm or at least a home. The adult male urgently needed a spouse as a home manager.

Even more questionable are data for the numbers of children born to the poor and to the rich. A child of the poor wasn't necessarily a burden to the family since he might easily have been able to earn his keep after the age of eight. Or the table could be turned around, showing the wealth of the adult male to be greater the more children he had, at least for a certain range. Nevertheless, one can say that those with wealth had almost twice as many children as the poor. By my standards, I choose to interpret this to mean that the poor were living with strong disadvantages. Some of these were related to housing and its distribution, a subject I highlight in this paper.

3.3 Dwelling-Derived Income

Are we willing to believe that relative inequality has remained roughly constant during the last two centuries? Is there other evidence about economic inequality for the entire population? I say yes, and that it appears dramatically in the value placed on living conditions of people, as measured by dwelling values. Surely house values directly represent the inequality of economic well-being in a most dramatic form and may, in a way, be an average measure of the total expenditure on food, clothing, shelter, and medical care, coupled with saving or dissaving.

3.3.1 United States in 1798

My preferred estimate of income distribution in 1798 for all families or individuals appears in figure 3.1 and in table 3.1. It is based on dwelling values and numbers of persons in houses, revealed in the 1798–1800 data sets. The distribution is lognormal in shape and demonstrates very substantial inequality, with a Gini coefficient of .6 or above, considerably more than that found for families and unrelated individuals at the end of the 1980s, when G was less than .5. There was a great range in the standard of living at that early time, as evidenced in the contrast between the home of Elias Haskett Derby, America's first millionaire, and the primitive log huts without windows or chimneys. There was large regional variation and, indeed, variation between townships and counties (see section 3.3.7). The standard of living, from this point of view, revealed extreme inequality, and averages or aggregates shown in national accounts seem far removed from reality for most individuals.

3.3.2 Utah in 1857

If only we had an expression for inequality for 1850 or 1860 that would allow us to see if long-run changes were taking place. One source of income distribution for all persons, or a large portion of the population, not reviewed in this paper derives from the early assessment records of the Mormon church. A study of income distribution among Mormon families in Utah in 1857 shows the share of total income of the lowest 40 percent of families to be 16 percent of aggregate income. This 16 percent portion is obviously larger than the 6 percent share for 1798. Yet it is not immensely larger, considering the fact that Mormon society at the time was relatively homogeneous. The Mormon distribution plots as a rough straight line on log paper and reflects the fact that some people were reported to be sickly or suffering from flooding or droughts.[4] The homogeneity of the Mormon data and the relatively late settlement date are definite limitations. We must turn to housing data for New York State to measure inequality change.

3.3.3 New York in 1798, 1875, and 1980

I will present distributions of dwelling values in New York State for 1798, 1875, and 1980—configurations roughly a century apart from each other—as a test of changing inequality. Admittedly, the data suffer from the fact that more than one family or unrelated individual might have lived in a dwelling, particularly in an urban setting. The fitting or "crowding" of individuals into dwellings is a complex affair. Some adjustments are made for the crowding of individuals by applying an elasticity coefficient to house value, but the issue is far from being resolved. Yet the data for 1798 and 1875 are particularly fruitful since the numbers of dwellings per adult male were similar in those years. Today there are almost as many dwelling units as there are adult males.

Again, I will focus on the shares of lower groups. Did the share of aggregate house value for the lower 40 percent of house values decline in the nineteenth century as lower groups suffered relatively? Or did our economy improve, in some fashion, in terms of the provisions transferred to the relative poor?

4. See Soltow and May (1979, 157). More elaborate distributions are presented by Kearl and Pope (1986, 222) and Pope (1989, 162); in a letter to the author, dated 27 March 1990, Pope discusses the problems inherent in the use of Mormon data to establish income distributions.

The 1857 Mormon array is the only complete nineteenth-century income distribution I have at my disposal for testing of lognormality. Its plot at percentiles 10, 20, . . . , 90, 95, 98, and 99 yields an excellent straight line on lognormal paper:

$$\log \text{income} = \underset{(.02)}{11.05} + \underset{(.020)}{730Z},$$

where Z is the standard normal deviate, $R^2 = .992$, and $N = 13$ (standard errors in parentheses).

There is some question about the extent of coverage, and the incomes of at least a few boys are included. Yet there also are comments such as, "been sickly nearly all the time," and "crop destroyed by frost."

The Data

The distributions of table 3.6 for 1798 are derived from the inventory of real estate made in New York and, indeed, in all states in that year. The table data say that the overall Gini coefficient for New York dwelling values was .74 in 1798, .66 in 1875, and .37 in 1980. These values can be amended by using continuity factors, but essentially they tell us that inequality decreased a little during the first century, and much more dramatically in the second.[5] Shares of lower groups rose relative to shares of the rich as the average value of a dwelling climbed, in current value, from $360 in 1798, to $3,500 in 1875, and to $37,000 in 1980, tenfold and then fivefold increases when adjusted with a consumer price index (David and Solar 1977, 16–17).

What did all of this mean to a family or individual at the fortieth percentile? The family's well-being, in housing, was enhanced materially in terms of flooring, ceilings, fireplaces, windows, and doors. The demise of log houses, as dramatic as the edifices seem to us today (and even as they appear in historical or fine arts museums), was a genuine step toward improvement. Log houses were the overwhelming construction type in 1798; the number of log houses recorded in the New York State censuses of 1855, 1865, and 1875 were 33,000, 20,000, and 13,000. These simple dwellings, having values that averaged 5 to 10 percent of the values of framed houses, rapidly disappeared. The family at the fortieth percentile had moved from a log cabin to a framed house.

Such must have been the story for many consumer items. The styles and qualities of clothing became less distinguishable. Homespun and linsey-woolsey disappeared as did log cabins. Stated more positively, our fortieth-percentile family probably owned a watch or clock in 1875, but not in 1798. Changes in transportation would be more difficult to quantify. The family probably was less likely to own a horse; it could, on the other hand, afford to buy a railroad ticket.[6]

The change in the housing situation in New York State is better understood by examining overall changes than by studying separately the changes in its urban and rural sectors. I attempt a classification based on New York and Kings counties, the only areas among the 1798 tax districts I could deem to be essentially urban. Table 3.6 shows little change in dwelling-value inequality in these two counties, but the issue is far from resolved; there simply is insufficient detail above the median in 1875, as presented in the table and in the discussion in its source note.

The dwelling distributions for counties other than New York and Kings

5. A case can be made for almost no change in inequality from 1798 to 1875, as stated in the note to table 3.6.

6. For clothing, see Kidwell and Christman (1974); Soltow (1990) includes a table showing that the distributions of timepieces in Scotland and Connecticut in 1798 were similar. Soltow (1981, 206, 210) deals with horse ownership.

Table 3.6 Distribution of Dwelling Values in North York State, New York and Kings Counties, and the Other Counties in 1798, 1875, and 1980

Class Limits ($)	All	New York and Kings	Other Counties
	Number of Dwellings in 1798		
30,001 and up	1	1	0
20,001–30,000	4	2	2
15,001–20,000	3	2	1
10,001–15,000	45	42	3
6,001–10,000	340	320	20
3,001–6,000	1,180	1,052	128
1,001–3,000	4,561	3,237	1,324
501–1,000	5,640	2,014	3,626
101–500	21,413	560	20,853
1–100	39,993	24	39,969
Total	73,180	7,254	65,926
Mean ($)	363	2,049	178
Inequality (*G*)	.737	.406	.626
Lognormal chart			
R^2	.978	.992	.997
Slope (*b*)	1.337	0.839	1.368
Aggregate value ($ millions)	26.578	14.863	11.714
Males twenty-one and up	118,000	13,910	104,090
	Number of Dwellings in 1875		
10,000 and up	65,373	53,060	12,313
5,000–9,900	60,942	33,272	27,670
2,000–4,999	144,311	23,604	120,707
1,000–1,999	151,310	5,429	145,881
250–999	201,357	2,126	199,231
100–249	66,326	727	65,599
50–99	14,286	116	14,170
10–49	7,134	24	7,110
Total	711,039	118,358	592,681
Mean ($)	3,300	11,720	1,590
Inequality (*G*)	.66	.38–.44	.59
Lognormal chart			
R^2	.999	—	.997
Slope (*b*)	1.428	0.71–0.83	1.219
Aggregate value ($ millions)	2,460	1,362	1,097
Males twenty-one and up	1,267,000	410,000	857,000
	Number of Dwellings in 1980		
200,000 and up	24,667	7,085	17,582
100,000–199,999	127,264	10,183	117,081
50,000–99,999	913,233	121,885	791,348
40,000–49,999	557,256	79,920	477,336
30,000–39,999	825,674	159,700	665,974
20,000–29,999	1,333,726	394,124	939,602

Table 3.6 (continued)

Class Limits ($)	All	New York and Kings	Other Counties
	Number of Dwellings in 1980		
10,000–19,999	1,310,429	479,706	830,723
5,000–9,999	259,874	93,714	166,160
2,000–4,999	24,267	8,944	15,323
Total	5,376,390	1,355,261	4,021,129
Mean ($)	36,520	29,080	39,030
Inequality (*G*)	.373	.371	.364
Lognormal chart			
R^2	.997	.975	.999
Slope (*b*)	0.69	0.68	0.68
Aggregate value ($ billions)	196	39	157
Males twenty-one and up	5,478,887	1,169,026	4,309,861

Sources: Soltow (1989a, 80); New York (1875, table 47); U.S. Bureau of the Census (1980, part 34, New York, 99, 537–42). Rentals or units were assumed to be 12 percent of dwelling values.

Note: A continuity adjustment factor can be applied within each class of the New York tables. The lognormal linear model can be fitted to the logarithms of the lower- and upper-class limits and their probits for each class; the resulting standard deviation of each class can then be translated into its interval Gini coefficient. These considerations raise the overall Gini coefficient for all, New York and Kings, and the other counties in 1798 from, respectively, .737, .406, and .626 to .769, .407, and .680. The 1875 distributions, adjusted upward by using these procedures, yield results very similar to those in 1798. In fact, one can make the case that housing inequality in New York State was very similar in the years 1798 and 1875.

demonstrate inequality decreases similar to those for the state, but they are less dramatic. In fact, one can argue that the measured drop in inequality for this group is rather small from 1798 to 1875: .626 to .590 for *G*, and 1.37 to 1.22 for the slope of the chart line. I tentatively conclude that this is my best evidence (aside from the issue of crowding) of change in income inequality in the nineteenth century.

Why did these distributions demonstrate so much inequality in the past? There were very large differences in standards from one county to the next. When this area variation is eliminated from the sixty-three tax districts in 1798, the inequality decreases dramatically.

The Dwelling House Value of the i*th Individual in the* j*th Tax District*	*All*	*New York and Kings*	*The Other Counties*
$G(DHV_{i,j})$	.737	.406	.626
$G(DHV_{i,j}\sqrt{DHV_j})$	.574	.372	.592

There simply were large area differences in the quality of housing in the early federal period in New York State.

Crowding

An obvious point of concern in making the comparisons in table 3.6 derives from the fact that there were 1.61 adult males per house in 1798 and that the ratio was 1.73 in 1875, and 1.02 in 1980. If housing values are to represent income distribution properly, then I should squeeze or "crowd" the men into houses in the earlier years to devise a house value per man (or perhaps per family). If 1.6 men were placed in each house or, say ten in every six houses in 1798, and it were assumed that there was equality in effective rental within each house, then relative inequality would remain roughly the same as stated in the table for houses. But what if more persons were squeezed into houses of higher (lower) value? Relative inequality would be thought to be less (more) than it now is for houses.

In the absence of concrete information for New York crowding, I must direct my attention to two most powerful theorems formulated by Aitchison and Brown (1969, theorems 2.6 and 2.3) concerning the jth moment and the reproductive property of a lognormal curve.[7] In reality, the theorems state that a methodical crowding based on dwelling value (V), say V^j or V^{-j}, decreases or increases the Gini coefficient some, but retains the lognormal form if the original distribution is lognormal. In the case of our distribution for 1798, a new class frequency $f_2 = f_1 V^{+.085}$, or $f_2 = f_1 V^{-.085}$, can effectively raise the total frequency about 60 percent; it alters the mean, lowers or raises the slope of the straight line on lognormal probability paper, and lowers or raises the Gini coefficient about 6–7 percent. The theorems state, in effect, that a frequency table with class frequencies and class means (f_i, V_i) ($i = 1, 2, \ldots, n$ classes) and which is lognormal in shape may be transformed into another lognormal curve by methodically squeezing (in a constant elasticity form) people into homes of larger values in a fashion lowering inequality some, but not much. Alternatively, squeezing more people into houses of lower value strengthens inequality. This alteration can be checked roughly by computer. Given the frequency table, with class frequencies and class means (f_i, V_i) for ten classes and its somewhat lognormal shape, one can obtain the new table with $f_2 = fV^{-.085}$, with each person having a value $V_2 = f_2 /(fV^{-.085})$; this raises the Gini coefficient from, say, .600 to .638.

This technical appeal is made to argue that the New York dwelling distribution probably almost represents rental value or income in all three periods. Inequality of rentals within the house can raise dispersion at least a little. Statistical evidence demonstrates that a distribution of dwelling values and its actual rentals had essentially the same relative dispersion.

3.3.4 Summary of Inequality Change

There was very extensive variation in the values of dwelling houses in New York State in the nineteenth century. Consider one spectacular example from

7. Theorem 2.6 states, "The jth moment distribution of a Λ-distribution (lognormal) with parameters μ and σ^2 is also a Λ-distribution with parameters $\mu + j\sigma^2$ and σ^2, respectively."

the 1875 census for Buffalo, where the Tenth Ward was a district for the wealthy and the Thirteenth for the poor. In the Tenth, there appeared a house valued at $500,000; it was a palatial mansion built between 1868 and 1870 by William Fargo, president of the American Express Company. The household, to be fair, included Fargo, his wife, daughter, two granddaughters, a second young family, and the butler, cook, parlor maid, chambermaid, nurse, and governor. In the Thirteenth lived John Madigan, laborer, and his family, in a house valued at $50.

Admittedly, these are extremes, so we should look at the share of total value of the lowest 40 percent in the array of dwelling values.[8] This 40 percent group, eliminating New York and Kings counties, accounted for 7 percent of value in 1798, 8 percent in 1875, and 18 percent in 1980; the maximum share for this group would have been 40 percent. One can say that the lower groups rose a little, from 7/40 to 8/40, a slight increase, and then spectacularly, to 18/40 of what might have been achieved with perfect equality.

3.3.5 Rents

All of my housing distributions are subject to the criticism that they do not properly account for the number of families or, say, the number of adult males living in a house. Even if this "crowding" phenomenon is unrelated to the value of the house, some consideration should be given to the shares of different persons in the house. Surely space and room values usually are not shared equally.

What I really would like is a distribution of *rents*, actual and implicit, for different families and unrelated individuals living in houses. The only early distributions I know about in this connection are those for the end of the eighteenth century in Amsterdam and Antwerp. These can be compared to the housing distribution in Boston where at least the number of persons in each house was enumerated. The Amsterdam data also allow us to measure changes in rent inequality from 1809 to 1914, an outside check on the fact that change did occur from the nineteenth to the twentieth century.

A special tax was assessed in Amsterdam in 1805.[9] It demanded not the ordinary inventory of house values as such, but an assessment of value for each person in the house, where the individual was either the major dweller or individual living in a room or in a half or complete cellar. My wife and I drew a sample from this inventory and found, much to my satisfaction, that the distribution of rentals was similar to that for dwelling values.[10]

8. The tabulated distributions for 1875 include at least a few hospitals, orphanages, buildings for religious orders, some large boardinghouses, and even some business or partial business edifices in which people lived. I am currently sampling manuscripts from urban areas in 1875, and have a rough estimate that about 10 percent of value shown for urban areas in 1875 should be eliminated. I do not think that inequality coefficients would decrease very much with this correction.

9. Details of Amsterdam housing data are given in Lievense (n.d., especially 49).

10. Sample drawn from Gemeente Archief, Amsterdam, archive number GAA 5045A, vols. 1–60; missing are vols. 2, 9, 11–13, 22, 23, 25, 27–29, 34, 53, 54. We recorded all units from

Did I know enough about economic distribution to state, before seeing the figures, that Amsterdam's frequency table would demonstrate more or less inequality than was true for Boston or for urban America in 1798? Certainly I thought that European cities in general, with their established hierarchies, should display more inequality. But I was warned by an Amsterdam archivist that there was an absence of hierarchy in his city. There were essentially no titled people, even by Danish or Swedish standards, let alone by those of France or England. There was no nobility, and Holland's government was seated in the Hague. Yet my working hypothesis had to be that there was more inequality in Amsterdam.

The distributions shown in table 3.7 do confirm my hypothesis, but not overwhelmingly, to say the least. Disparities appear more at the top than among the lower groups. Consider the shares of those with rental below the median, the complements of the upper shares stated in the table: Amsterdam, 1805, .132; Antwerp, 1799, .133; Boston, 1798, .160; U.S. urban, 1798, .112. Boston showed less relative deprivation, but some of the differential is due to measurement error. I had to assume equality of rentals among dwellers within housing units in Boston. To have made this equality assumption for Amsterdam would have almost eliminated the differential. Thus the startling conclusion must be that lower shares were very little lower in Amsterdam. This statement surely does not apply to either Copenhagen, Stockholm, or London. And the Amsterdam-Boston results do not include the influence of country estates owned by the rich.

Some mention should be made of the results in table 3.7 that demonstrate that inequality in Amsterdam's distribution of housing in the twentieth century was less than it was in the nineteenth century. These results parallel those found earlier for the distributions in New York State. I have found similar results in another context from a study of housing in Scotland.[11]

3.3.6 Poor in Ontario Shanties

For statistical purposes, how convenient it would be if there had been a census each year in the early part of the nineteenth century, enumerating in detail those poor individuals or families in the lowest decile range, in the

every tenth page of vols. 1, 4, 8, 16, 20, 24, 32, 36, 40, 44, 48, 52, 56, and 60, and all pages from these volumes outside the selected interval from f. 31 to f. 799. The sample was thus from 154 pages to obtain 1,062 units from within the selected interval and 1,487 pages to obtain the 1,884 units outside the interval. These counts exclude the 3.6 percent of units that were not rented. Several dozen dwelling units were in warehouses, piers, and businesses. I eliminated one *grondhuis* with very nominal rent. There were 4,681 pages in all of the extant volumes.

Aggregates of the verponding evaluation in each of the 60 *wijks* (GAA 5045) in 1785–87 show that the *wijks* missing for 1805 represented 22.45 percent of value. This factor and the page-count factors stated above indicate that the 2,946 observations in the sample represented 6.0 percent of the 49,300 units in the city at the time.

11. Soltow (1971). This study was derived from the censuses of housing in Scotland since 1861; these censuses encompass all persons in housings units, not just the rich.

Table 3.7 **The Relative Distribution of Annual Rental Values in Amsterdam in 1805, Antwerp in 1799, Boston and Urban Areas in the United States in 1798, and Amsterdam in 1919**

	A_R, the Proportion of Aggregate Rental Value of the N_R Group				
	Amsterdam 1805	Antwerp 1799	Boston 1798	U.S. Urban Areas 1798	Amsterdam 1919
Proportion of all cases (N_R)					
.01 (top)	.102	.107	.091	.109	.053
.02	.166	.183	.149	.174	.105
.05	.297	.335	.266	.311	.224
.10	.453	.477	.396	.456	.337
.20	.644	.649	.566	.638	.482
.50	.868	.867	.840	.888	.742
.80	.965	.970	.964	.982	.921
.90	.987	.990	.986	.995	.967
1.00	1.000	1.000	1.000	1.000	1.000
Mean	f160	f194	$131	$77	f142
Inequality (G)	.581	.591	.520	.602	.386
Number of units	49,300	13,800	4,245	73,000	141,556
Number in sample	2,946	1,376	2,423	6,780	141,556
Population	200,000	51,000	23,000	320,000	640,000
Units/population	0.25	0.27	0.18	0.23	0.22
Straight-line model plotted on a lognormal chart					
Correlation (R^2)	.976	.981	.994	.994	.953
Slope (b)	0.949	1.047	0.939	1.275	0.567

Source: For Amsterdam in 1805, see Soltow (1987b, table 6). I took a sample of every tenth item from Antwerp (1799), chosen because the listing of rental values within each building was more methodical than in 1796 or 1797; see also De Belder (1977, 3–4); Lis (1986, 71, 76); Reyniers-Defourny (1979). Data for Boston and the United States appear in Soltow (1987a, 1989a). Data for Amsterdam 1919 are from Amsterdam (1923, 19).

lowest quintile range, and below the median and fortieth percentiles. This model census would have stated the average income or other characteristics of individuals in each decile range. Even the magnificent quinquennial censuses of Sweden from 1805 to 1855 tell us only the numbers of those deemed destitute, poor, somewhat rich, and rich. I search for any tallies of lower income subsets in the century that can serve as proxies for any one of these ideal census measures that cover a significant portion of the population—counts covering more than the few percent who were orphans, blind, hospitalized for mental illness, and supported widows.

One possibility is to examine those families or households living in "shanties," as enumerated in the Ontario, or upper Canada, census of 1851. The houses were classified and ordered in such a way that they suggest a skewed, if not a lognormal, distribution, as shown in table 3.8.[12] Values were not stated, but those reported for stone, brick, frame, and log for New York in 1855 coupled with Ontario frequencies provide some insight by suggesting a lognormal curve with a Gini coefficient of about .7 and a shanty value of $5 to $20.

A study by regions shows large variation, given in table 3.9. Those counties with greater development were less likely to have poorer housing. Thus there was an inverse relationship between the proportion of shanties and the proportion of land held that was cultivated, as demonstrated in the county classifications.

What were the characteristics of persons living in shanties? Were they beginning settlers? Were they young, with relatively few children? We can examine some of the demographic characteristics by using the sample of 1,201 dwellings presented in table 3.10. Turning first to families living in shanties, we see that their average age was surprisingly high, being only a year or two less than the overall average. Occupational distribution was decidedly different from that of the total population; household heads living in shanties were more than two-and-one-half times as likely to be farmers or those with "other" occupations. Shanty heads were more likely to be foreign-born than might be expected, leading one to suspect that they may have been in Ontario for shorter portions of their adult lives. Finally, 42 percent were Roman Catholic as opposed to the 20 percent proportion for all heads. Shanty heads tended not to be Methodists or Baptists or members of minor sects.

Evidence from the large minority group of log-house heads reveals surprisingly few differences from the characteristics of all persons. Their ethnic or birth traits were representative; the one exception was that there were relatively fewer among them who were born in the United States. They tended more to be farmers, not laborers or those with other occupations.

A further sample was devised to furnish more suggestions about the demographic characteristics of persons living in shanties. A sample of every tenth dwelling from the principal sample yielded the data in table 3.11 on family composition. Judging from the age of the oldest Canadian-born child and the youngest foreign-born child for twelve shanty families, I find the median length of residence in Ontario to have been eleven years, not two to five years as I had expected. This evidence indicates that shanty families were not in earlier stages of their life cycles. They were much less likely to have been native-born and were more likely to have been Roman Catholics. These data could be developed much more generally. Results for foreign-born could be compared to those of immigrants to Canada in an earlier period, as suggested by Bernard Bailyn (1986, chaps. 6, 10).

12. Aitchison and Brown (1969, 27) suggest that classifications based on some homogeneity principle often are approximately lognormal.

Table 3.8 Dwelling Types in Ontario in 1851

Type	Frequency	Proportion	Value (New York, 1855, $)
Stone	4,211	.029	6,900
Brick	5,117	.035	5,500
Frame	53,931	.370	785
Log	65,503	.449	46
Shanty	17,191	.118	5(?)
Total	145,953	1.000	

Source: Canada (1851–52, appendix 15, table 8, 430–31); New York (1855, 249–50).

Table 3.9 The Proportion of Dwellings That Were Shanties (PSHAN) and Log or Shanty (PLOGSHAN) in Ontario for Forty-two Counties in 1851

PSHAN	Frequency	PLOGSHAN	Frequency
.40 and up	4	.80 and up	10
.20–.39	4	.60–.19	15
.10–.19	16	.40–.59	12
.05–.09	13	.20–.39	6
.01–.04	4		
Under .01	1		
Total	42		42

Source: Canada (1851–52, 402–29).

3.3.7 Poverty Areas in the United States in 1798

A promising field for the historical investigation of the degree of well-being of persons below median income can be derived from studies of many areas such as townships, counties, states, or larger units. Usually there is large variation in shares of the poor when using these area classifications, but the shares can be related to other economic variables of the localities. Detailed studies of this type demand very large data sets indeed; examples of such are the 165,000 property values in Ohio in 1835 and the distribution of housing within each of 1,200 townships in New York in 1875.

The possibility of a historical study of areas exists for 687 tax districts in the United States in the year 1798 even though data are not as complete as one would like, particularly for southern states. What index can be used as an indicator of relative poverty? The fertility of the land and its relative distance from urban areas or ports can be an element in the ability of an area to provide jobs and land for lower groups. I consider the average value of land per acre (VAC) for each of the 687 to be an expression of these characteristics. I plotted the 687 VAC and colored those where VAC was $5 or more. I optimistically hoped that well-defined contours would in some way tie into Bailyn's

Table 3.10 Characteristics of a Sample of 1,201 Dwellings in Ontario in 1851, Classified by Type of House

	Shanty	Log	Frame	Stone, Brick	Other	All	Rural	Urban
Sample size	146	578	403	60	14	1,201	1,140	61
Age, household head	41	43	42	48	41	42.9	43	42
Stories	1.0	1.03	1.2	1.4	1.1	1.123	1.1	1.3
				Proportion of Cases				
Occupation, household head								
Farmer	.45	.73	.40	.58	.50	.574	.60	.16
Laborer	.32	.10	.10	.05	.07	.126	.12	.15
Other occupation	.17	.12	.44	.37	.43	.251	.23	.62
Female	.05	.03	.04	.00	.00	.033	.03	.07
None, unemployed	.01	.02	.02	.00	.00	.017	.02	.00
Total	1.00	1.00	1.00	1.00	1.00	1.000	1.00	1.00
Birth, household head								
Foreign-born	.88	.75	.68	.73	.79	.742	.74	.84
England	.08	.10	.19	.15	.29	.133	.13	.13
Scotland	.15	.19	.10	.20	.14	.157	.15	.20
Ireland	.62	.33	.21	.18	.29	.316	.31	.48
Germany	.00	.05	.03	.03	.00	.036	.04	.00
United States	.03	.07	.13	.13	.07	.089	.09	.02
Canadian-born	.12	.25	.32	.27	.21	.258	.26	.16
Ontario	.08	.22	.29	.27	.14	.231	.24	.13
Quebec	.03	.03	.01	.00	.00	.019	.02	.02
Religion								
Catholic	.42	.22	.10	.08	.00	.197	.19	.33
Baptist	.02	.04	.04	.05	.00	.037	.04	.02
Church of England	.24	.17	.25	.25	.29	.211	.21	.25
Methodist	.08	.16	.27	.23	.03	.188	.19	.15
Presbyterian	.21	.27	.15	.23	.14	.217	.22	.21
Other	.03	.13	.20	.15	.50	.146	.15	.05
Total	1.00	1.00	1.00	1.00	1.00	1.000	1.00	1.00

Source: Canada (1851–52). I recorded all entries from each of about 330 pages. My procedure considered even-numbered films and every eightieth page from this personal census.

Table 3.11

	Shanties	Nonshanties
Number	15	97
Proportion with wives	.83	.82
Average number of children	3.51	3.54
Average number of sons twenty and older	0.3	0.33
Median number of years in Ontario	11	Not computed

(1986) beautiful immigration maps, but I must candidly admit that this project failed. First, I had difficulty in obtaining precise latitudes and longitudes; even plottings for each of the 359 counties proved difficult because of strong variation in VAC between contiguous counties. Collapsing contiguous areas into 180 plots did demonstrate concentration of fertility areas in the Shenandoah Valley, eastern Pennsylvania, and southern New England. Yet in itself such a map really does not offer much of any explanation of relative inequality among *persons*. Areas in which VAC was under $5 very well may have been those with more acreage per farm. Even the average value of land per adult male tells mc little about distribution within each district.

A more direct measure of shares of lower groups in 1798 can be determined from stated or estimated proportions of houses in a district that were under $100 in value. Even here there is doubt, since districts either somewhat or far removed from urban centers had both lower prices and a large majority of houses in this category. It seems better to express dollar dwelling values as a proportion of the mean value in order to consider relative shares. To this end I will focus on the share of aggregate dwelling value below the fortieth percentile, SDHV40. It is the aggregate value of the poorest 40 percent of houses, expressed as a proportion of the overall aggregate value; this measure may vary from .00 to .40.

A background display of the part this measure plays in the distribution of U.S. dwellings derives from a sample of 39,890 items and from stated frequencies (computed by authorities in 1798) for ten frequency classes of dwelling values. This study produces an essentially lognormal distribution both nationally and for each of seven areas, presented in figure 3.3. The regularity of plots adds confidence in stating lower dwelling-value shares in each of the seven districts, as given in table 3.12. In general, the shares of these "poor" houses were quite low—9 to 10 percent of aggregate value. Differences must be treated with caution, particularly those for the South, the region with less complete data. Within the North, shares interestingly were larger for rural areas nearer cities than those further removed either to the West or to the North. The poor in northern urban areas were almost as well-off as in rural areas, and better-off when more than eighty miles removed from the six major East Coast cities. Southern areas seemed to follow the same scheme. Rural southern areas further removed from the East were influenced by values in Tennessee and Kentucky.

Let me now further disentangle these shares by considering areas in finer detail. In general, the relative inequality of an entire large area is greater than the average relative inequality of its counties and greater yet than the average relative inequality of its more numerous tax divisions. The overall Gini coefficient for the rural North within eighty miles of its major cities is larger than the average coefficient for its fifty-one counties, and this, in turn, is larger than the average for its 233 tax districts. The overall coefficient is composed of inequality within districts and that between the means of its districts. The

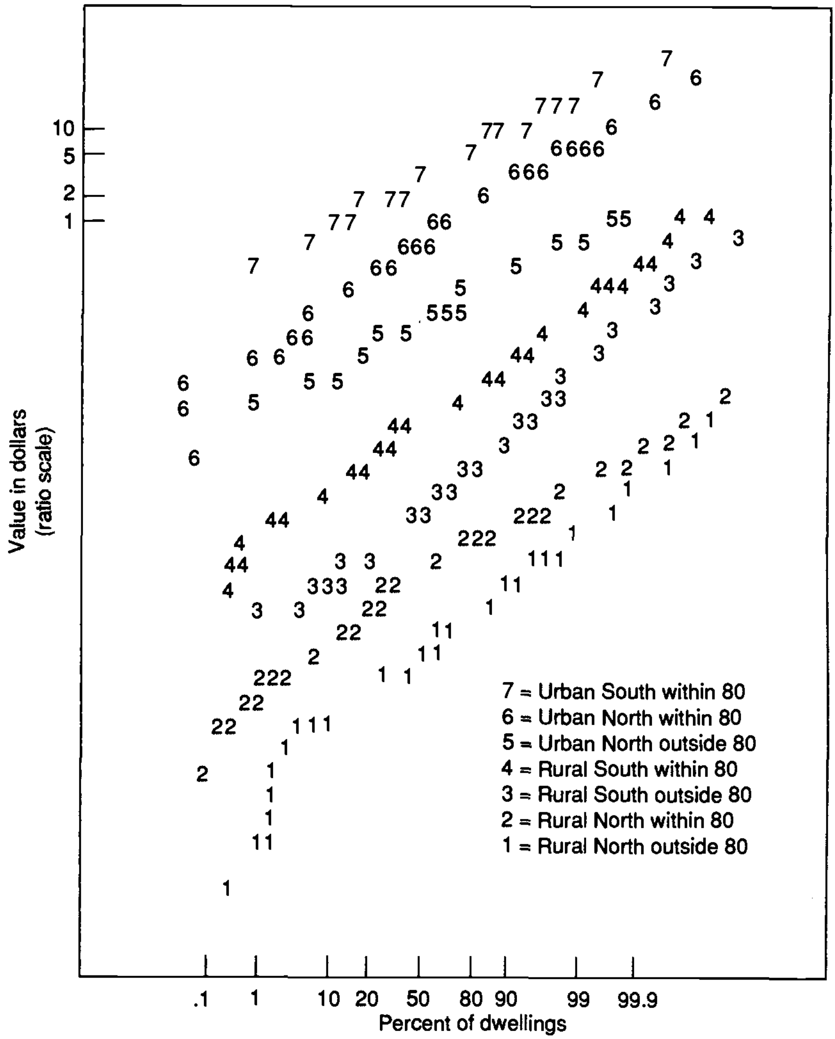

Fig. 3.3 A tier chart of the distribution of values of dwellings in each of seven regions in the United States in 1798
Source: See table 3.12.

shares of the lowest 40 percent presented in two different columns of table 3.12 reflect these considerations. Also, the columns reflect the fact that there was more detail for some regions than for others. Yet it is fruitful to explore this analysis further.

The estimate of SDHV40 for each of the 687 districts in the United States demands some kind of interpolation from among class limits and class frequencies of ten tax-rate classes. I have derived estimates by applying the log-

Table 3.12 The Proportion of the Aggregate Value of Dwellings Accounted For by Persons below the Fortieth Percentile (SDHV40) in Seven Sections of the United States in 1798

	Considering All Values within a Region	Considering All Values within Each Tax District	
Seven Regions	SDHV40	Number of Districts	Average SDHV40
Rural north	.091	233	.101
Within 80 miles			
Outside 80 miles	.069	153	.090
Urban north	.082	44	.153
Within 80 miles			
Outside 80 miles	.109	2	.135
Rural south	.064	70	.088
Within 80 miles			
Outside 80 miles	.036	75	.066
Urban south	.106	4	.141
Within 80 miles			
All	.0916	681	.101

Source: SDHV40 is computed from the seven distributions of housing values derived from a sample of 39,890 dwelling values and stated frequency counts for ten classes, as described in Soltow (1989a, chaps. 2, 3, and appendixes 1, 5). The SDHV40 for each of 681 tax districts was computed from lognormal equations applied to estimates of district Gini coefficients (Soltow 1989a, 80 n. 20). Shares for six districts could not be determined.

Note: Whether the district was within eighty miles was determined by the minimum distance between a district and one of the six major cities: Boston, New York, Philadelphia, Baltimore, Norfolk, or Charleston.

normal considerations used previously within each of the ten in developing 687 Gini coefficients. From each of these I have computed SDHV40 assuming overall lognormality within each district. The shares of the lower 40 percent of houses varies quite substantially from district to district, as shown in table 3.13. The distribution is at least slightly skewed and reminds one in a superficial sense from the standpoint of economics, but in an engaging statistical sense, of the distribution by county of pauper rates for England and Wales in 1803, a distribution with about the same average and relative dispersion.[13] But surely persons living in poorer houses in the United States were enjoying far better economic conditions.

What are the regional variation characteristics of the 40 percent shares? Again, I had hoped to present a map by plotting shares at their respective latitudes and longitudes, but the variation in shares among contiguous tax

13. Great Britain (1805, 13: 714). A clearer analogy could be made by studying the proportion of houses under, say, $33 in rural areas and under $100 in urban areas of the United States in 1798. For a plot of the distribution of pauperism rates, see Yule (1919, 92–93).

Table 3.13 Value Shares of Poorer Houses (SDHV40) for Each of 681 Tax Districts in the United States in 1798

SDHV40	Number of Districts
.25 and up	3
.20–.249	9
.15–.199	81
.10–.149	253
.05–.099	236
.02–.049	84
.007–.019	15
Mean	.101
Gini coefficient (*G*) of SDHV40	.254

Source: See table 3.12. Transformation equations giving shares from Gini coefficients are stated in Aitchison and Brown (1969, chap. 2).

Note: Using the terminology of the Statistical Analysis System and the Gini coefficient, G, for each district, I have σ = ((2**.5) * PROBIT(1+G)/2), and then, SDHV40 = PROBNORM(PROBIT(.40) − σ). It proved to be impossible to compute G for six districts reporting frequency totals only for the lowest tax class.

districts made such a procedure very cumbersome. Consider instead a few regression equations. There was a tendency for shares to be larger in districts with *higher* dollar land value per adult free male (VALAND21).

$$\text{SDHV40} = .089 + \underset{(.000005)}{.000020}\ \text{VALAND21};$$

R^2 = .020, N = .681. This persisted to a certain extent if the minimum number of miles from one of the six major coastal cities (MILES6) is considered.

$$\text{SDHV40} = .123 + \underset{(.000005)}{.000011}\ \text{VALAND21} - \underset{(.000019)}{.000281}\ \text{MILES6};$$

R^2 = .27, N = 681. The proportion of the lower 40 percent tended to rise .011 with each $1,000 increase in per capita wealth in land (excluding houses) and to go down .028 for districts one hundred miles from urban centers.

Another way of stating the pattern is to consider the age of the tax district by considering the formation date of its county. When this variable is added to the above equation, its regression coefficient shows a significantly larger share for the lower group in areas of older age. It is doubtful that all of this pattern could be a shanty effect or a log cabin effect, since the majority of dwellings in rural areas in both the East and the West were log dwellings.[14]

14. $$\text{SDHV40} = .599 + \underset{(.00004)}{.00012}\ \text{VALAND21} - \underset{(.00002)}{.00017}\ \text{MILES6} + \underset{(.0004)}{.0028}\ \text{AGE};$$

R^2 = .33, N = 681.

Surely it would be advantageous to have expressions for lower groups for just before the Civil War, or even just after, for the United States or its northern region. The effects of settlement and its changes then could be better understood. The data set just discussed, that of housing in 1798, is probably the most important among the nine sets presented in this paper. It gives detail for the entire country and for its various areas, and gives values for those below the median, something not available for the sets dealing with real estate values, wealth, and census farms. Yet it is inferior to data for stature or for New York housing in the sense that we have no feeling for changes in shares occurring in various regions of the United States over long periods of time.

3.4 Inequality in Height and in Nutrition

Any variable that registers a value for each and every member of a population must be attractive to anyone wishing to study shapes of distributions for economic characteristics of individuals. The height or stature of adult males continues to be a prominent variable that measures, at least in part, the well-being of a large segment of society's labor force. But what statistical procedures should be applied if one wishes to gauge economic changes? Robert Fogel (1986, 456) states that the genetic and environmental components of height are difficult to assay, but "for most well-fed contemporary populations, . . . systematic genetic influences appear to have very little impact on *mean* [my emphasis] heights." Margo and Steckel (1983, 168) assert that "although genes are important . . . physiologists and nutritionists agree that differences in mean height across populations are largely the result of environmental factors." In another article, Steckel (1983, 3) again notes the role of environment in average height differences, but states that "genes are important determinants of the heights of individuals." These men, and others, often employ procedures involving regression analysis where the dependent variable is height, not height below the median, specific quartile, or decile in the array of heights.

Emphasis on means and multiple regression equations can be somewhat tangential in studies of dispersion in heights. This can arise even though investigators are very keenly aware of inferences that can be made about distributions. Thus, John Komlos (1990, 607) begins his most recent article on height with this statement, "The secular trend in the distribution of income and wealth . . . has been a topic of concern . . . [in studies dealing with] the last two centuries." Yet nowhere does he present a distribution of heights, let alone a Lorenz curve of heights of individuals.

There is some silence concerning the shapes of the distributions of heights, an unwillingness to present frequency curves of stature at different dates in a fashion where they can be compared at various percentiles. If we had plottings of Lorenz curves or probit curves of heights in 1812, 1864, 1918, and 1980, then we might judge the changes in the inequality in height itself.[15] Such an analysis could serve as a focus for the significance of measures of stature. If height, income, and nutritional adequacy are positively related to each other in, say, 1812 and again in 1863, then we should focus on the relative changes among the short and the tall in this half-century. This approach could provide insights into changes for both the poor and the rich.

One who is continually interested in the degree of equality or inequality within a society, and how much it might change, becomes frustrated if statements about inequality that appear in the literature of economics cannot be tested. An early example of such was made by Edgar Martin (1942, 57) when he asserted, "Nowhere was there greater contrast between the diet of the rich and that of the poor than in the South." In this respect, what do height data say? One might even blatently ask if the lowest 40 percent in stature had a smaller proportion of aggregate height in the North as compared to the South.

3.4.1 Height Distributions

I attempt to contrast inequality in height by means of a series of distributions for the United States, as presented in table 3.14, for periods encompassing the War of 1812, the Civil War, and the First World War. To be sure, one can argue that the figures are not comparable from the standpoint of being representative of age, ethnic and nativity mixture, color, or region. Yet the data do represent large samples of recruits, a group that perhaps often included those who soon would be rejected on account of physical defects. I use my sample for 1799–1819 that was originally drawn for a study of illiteracy, since I later will present illiteracy rates and their changes for the short and the tall from 1799 to 1894. In general, the Civil War heights exclude those from the South and thus may tend to understate slightly the degree of inequality. Yet an analysis of height inequality related to specific age, nativity, and residence regions, to be presented shortly, indicates that the matter is not important; data for 1799–1819 and 1918 for the North and South confirm this conclusion. The information for the Civil War definitely includes individuals before rejection and, according to Baxter (1875, 1: vii.), "may be said to represent the adult male population without selection." The height study for 1976–80 is a sample of males in general and is unrelated to the process of army recruitment.

The height distributions for the four periods given in table 3.14 are convenient since they differ in dates by about fifty to sixty years and cover, in a

15. The first rather elegant presentation of height as a normal curve was achieved by Elliott (1863, 41–44, diagram C), using a chart and equations, at the International Statistical Congress in Berlin in 1863. Surely Elliott would have compared his data set with the distribution for a later date had he had access to information such as that obtained by Davenport and Love (1921, 67–74) for the First World War. Van Wieringen (1979, 1986, 318–19) presents a table depicting height percentiles for 1950 to 1978 and 1983 for Dutch draftees; he also gives a chart that depicts cumulative frequencies above specified heights for years after 1850.

Table 3.14 **Nine Distributions of Heights among Army Recruits in the United States in 1799–1819, 1861–65, and 1918, and in a Sample of the General Population in 1976–80**

	1799–1819		Civil War Period			1918		1976–80	
Height (inches)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
75 and up	.0099	.0115	.0031	0151	.0202	.0056	.0044	.0471	.0506
73–74	.0261	.0300	.0226			.0262	.0236	.1344	.1323
71–72	.1110	.1228	.0989	.0645	.0808	.1028	.1003	.2665	.2673
69–70	.2444	.2561	.2306	.1772	.2046	.2291	.2342	.2812	.2789
67–68	.2863	.2830	.3052	.2887	.2993	.2980	.3084	.1801	.1826
65–66	.2145	.1962	.2308	.2636	.2413	.2161	.2154	.0699	.0664
63–64	.0716	.,0672	.0915	.1418	.1172	.0907	.0865	.0160	.0175
61–62	.0218	.0189	.0137	.0416	.0312	.0246	.0208	.0036	.0030
Below 61	.0144	.0143	.0036	.0075	.0054	.0069	.0064	.0012	.0014
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Mean	67.8	67.9	67.75	67.30	67.67	67.49	67.60	70.0	70.1
Inequality (G)	.02268	.02270	.021	.02219	.02197	.02236	.02166	.02121	.02129
Lognormal model									
N (classes)	19	19	16	8	8	21	21	17	17
R^2	.979	.980	.998	.999	.999	.996	.994	.987	.987
Slope (b)	0.045	0.045	0.037	0.039	0.039	0.042	0.041	0.041	0.042
G, sixteen classes[a]	.0227	.0227	.0211	.0221	.0219	.0223	.0216	.0216	.0217
Number of men	1,452	1,232	719,438	501,068	315,620	867,755	66,885	2,236	2,236

Sources: Distributions 1 and 2 are derived from a sample of 2,762 army enlistees drawn from Record Group 94 in the National Archives, Washington, D.C., for the years 1799–1894. Further details are given in Soltow and Stevens (1981, 52).

Distribution 3 is derived from sixty-eight distributions involving seventeen age classes for each of four nativity groups; each distribution has thirty height classes. The total number of recruits or enlistees accounted for in these classes is 729,320. See Gould (1979, 96–103). Gould presents a summary table of the distribution of ages of 1,012,273 enlisted volunteers during the Civil War on page 34 of this study; I standardized distributions 1, 2, and 9 so their relative ages from 18–35 or 18–34 reflect Gould's distribution in those age ranges.

Distributions 4 and 5 involve over half a million recruits (and those rejected) during the Civil War. The frequency tables have only eight height classes (59″, 61″, . . . , 73″) (Baxter 1875, 2: 2–81, 166–97).

Distributions 6 and 7 involve draft recruits in 1918, as stated in Davenport and Love (1921, vol. 15, *Statistics*, pt. 1, 109). Disbributions have twenty-one classes for height and twenty-two classes for nativity and ethnicity. I was unable to standardize for ages.

Distributions 8 and 9 are derived from two frequency tables, one for males 18–24 and the other for 25–34; each has twenty-one height classes. I weighted the two distributions so they would reflect Gould frequencies (United States 1987, 24, 35).

Notes: The distribution frequencies are proportion of total cases. The distributions are for a sample of army enlistees, 1799–1819, ages 18–35, age standardized, (1) all color, all nativities, (2) white native-born; recruits and draftees in the Civil War, (3) ages 18–35, birth in northern states and Ireland, (4) all colors, all nativities, age unspecified, (5) native-born white, age unspecified; draft recruits in 1918, (6) all colors, all nativities, (7) agricultural North, native-born whites over 73 percent; general population of males, 1976–80, (8) ages 18–34, (9) Ages 18–34, age adjusted (frequency as a proportion of total cases).

[a]The degree of inequality measured depends on the types of individuals included in the population and is particularly sensitive to age and nativity composition. Frequency tables with, say, eight classes instead of sixteen generally will have a smaller G because no allowance is made for inequality between the lower

Table 3.14 (notes, continued)

and upper class limits of a class. Finally, published tables with or -end classes demand some assum tions about their class means. These and other aspects of measurement mean that the results for both and *b* in this table must be treated with caution.

One possible procedure for handling the nine distributions uniformly is to consider the one-in classes 61, 62, . . . , 74, coupled with two open-end classes below 61 and over 75 with assumed mea of 60 and 76. These sixteen classes have been used as my preferred measure of inequality in this tabl It was necessary to make interpolations using the lognormal slope to obtain sixteen or seventeen class from the eight published classes in distributions 4 and 5. This procedure for establishing sixteen class provides the arrangements used for constructing the Lorenz curves in table 3.15; the 1799–1819 heigt 75 inches and taller average 76 inches, but include three men at 88 inches and one at 82 inches.

general sense, most of the last two centuries. They allow us to make statements about changes in inequality of stature over long periods; I shall make inferences about these changes because they may reflect changes in the distributions of income and wealth. To be realistic, one must realize that demographic and genetic influences may be dominant and may swamp economic influences. Changes in marriage pairings, differential fertility rates, and child birth-order patterns impose influences not accounted for in the table. Other factors such as age and nativity can be controlled to a certain extent.

I would like my distributions of army recruits to be representative of the population as a whole if I am to make inferences with respect to general economic conditions. If the foreign-born among adult males was 10 percent of the population in 1812 and 30 percent in 1864, I would like to apply these same proportions to all data for army recruits. Nevertheless, I would like to study dispersion for just white native-born and for specific age ranges, preferably with the same age mixture. Yet from a practical standpoint, these matters may not be of major consequence.

The Baxter data for more than a half million recruits during the Civil War provide a distribution of thirty height classes (from below 61 inches to above 75 inches) for each year of age from sixteen and under, to thirty-five and over, in the case of four nativity classes. These yield the following equation for those eighteen and older:

$$\text{Gini coefficient of height} \times 10^5 =$$

$$\underset{}{2{,}058} - \underset{(.67)}{1.98\ \text{AGE}} + \underset{(8.8)}{7.8\ \text{NB2}} + \underset{(8.8)}{2.7\ \text{NB3}} - \underset{(8.8)}{100.2\ \text{NB4}}$$

$R^2 = .80$, $N = 60$ cells, where NB1 = 1, if born in New England; NB2 = 1, if born in New York, New Jersey, and Pennsylvania; NB3 = 1, if born in Ohio and Indiana; and NB4 = 1, if born in Ireland. Gini coefficients for younger age groups were slightly larger. Inequality varied very little for regional nativity groups in the United States. Surprisingly, inequality among Irish-born was less—about 5 percent less, on the average. The data in table 3.14 for my sampling of 1799–1819 allow age weighting, which forces the same age distribution and average age as in 1861–65, and allows separate representation of native-born.[16]

Each distribution in the table is essentially lognormal in shape. The R^2 coefficients are often greater than .990, but did fall to as low as .980 or lower when detail allowed the plotting of points in the upper and lower tails beyond two or three standard deviations. A tier chart of the nine distributions given in table 3.14 shows essentially nine straight lines parallel to each other (with slopes, *b*, roughly the same), demonstrating the very surprising fact that relative inequality remained essentially constant for the period of nearly two centuries under consideration. The Gini coefficients are also about the same for each of the distributions and present an alternative verification of the relative constancy of inequality.

To better understand the results of table 3.14, let me consider first the initial distribution. It says that the degree of inequality for 1799–1819 was .02268, as registered by using the Gini coefficient. The distribution plots as an excellent straight line on lognormal probability paper, with

$$\ln \text{height} = 4.22 + \underset{(.0001)}{.040Z},$$

$R^2 = .98$, fitted to 1,452 sample points (Z is the standard normal deviate). The slope 0.040 is an estimate of the standard deviation of the logarithms of heights. Only one point, that for an individual 82 inches tall, deviates appreciably from the line. These results are conservative estimates of inequality in the sense that heights under 60 inches did not appear, and those of boys under eighteen years have been excluded.

3.4.2 Lorenz Curves of Height

The Gini coefficient of height, $G(H) = .02268$, is quite small, at least relative to that for income distribution of the population as a whole at the time, perhaps as high as .60, as suggested by considerations of dwelling income in 1798 (Soltow 1989a, 247.) A lognormal distribution with $G(X) = .02$ can be transformed to one with $G(X^{33}) = .60$, using the convenient exponent of 33; I propose employing this magnification factor with heights, especially in comparing relative shares used in plotting Lorenz curves.[17] Inequality in the first 1799–1819 distribution becomes $G(H^{33}) = .64$, as determined from a computer run and as stated in table 3.15. Such a procedure conceivably implies, in a dangerous fashion, that a height distribution could be transformed into an income distribution using an elasticity of 33, a most unlikely possibility in the absence of data for other periods for income or height.

16. Age-specific distributions for white and colored native-born recruits in the United States during the Civil War show the former group to have less relative inequality. For men accepted, see Baxter (1875, 2: 199–299).

17. The factor of 33 is derived using theorem 2.1, found in Aitchison and Brown (1969).

Table 3.15 **Lorenz Curve Shares of Adjusted Aggregate Height, H^{33}, of Army Enlistees and Others, for the Nine frequency Tables from 1799 to 1980 Presented in Table 3.14**

	$A(H^{33})$, the Proportion of Aggregate Height of the $N(H^{33})$ Group								
	1799–1819		1861–65			1918		1976–80	
$N(H^{33})$, the Top Proportion of All Men	All 18–35, A.S. (1)	Native White 18–35, A.S. (2)	All Irish-Born 18–35 North (3)	All (4)	Native White, A.S. (5)	All (6)	Native White 73% of Agricultural, North (7)	General Population 18–34 (8)	General Population 18–34, A.S. (9)
.01(top)	.188	.173	.136	.159	.164	.158	.149	.069	.067
.02	.260	.260	.197	.227	.234	.226	.212	.137	.134
.05	.385	.385	.325	.363	.367	.357	.341	.331	.336
.10	.508	.508	.468	.503	.508	.495	.479	.470	.477
.20	.668	.667	.642	.670	.671	.665	.650	.639	.644
.30	.779	.780	.754	.777	.779	.771	.758	.754	.757
.40	.855	.851	.832	.851	.849	.846	.836	.828	.830
.50	.903	.901	.890	.900	.898	.900	.892	.886	.888
.60	.939	.939	.929	.938	.939	.936	.931	.931	.933
.70	.966	.966	.960	.966	.965	.965	.962	.960	.961
.80	.984	.983	.981	.984	.983	.984	.982	.981	.981
.90	.995	.996	.994	.995	.995	.995	.994	.994	.994
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$G(H^{33})$	.645	.642	.610	.640	.638	.634	.619	.602	.606

Source: See table 3.14; results must be treated with caution since an extreme height has a significant impact. For an early statement of this problem, see Sheppard (1897).

Note: Adjusting height, H, uses an elasticity coefficient of 33 for the sixteen classes and frequencies stated in table 3.14. A.S. = age standardized.

What is the evidence of height inequality in other periods? Tables 3.14 and 3.15 and the Lorenz curves in figure 3.4 show the surprising general result that inequality in height changed very little in the four periods.[18] Measurement error arising from the choice of lower- and upper-class limits of frequency distributions and midpoint assumptions might alone account for stated differences in Gini coefficients and also in lognormal slopes. The fact that Civil War figures largely exclude those from southern states may reduce the Gini coefficient by 2 to 3 percent.[19] It is particularly true that the drop in the Gini coefficient in this century, shown in table 3.14 to be about 5–10 percent, might be either larger or smaller if, say, Vietnam draftee heights were employed. Perhaps it goes too far afield to discuss changes in height inequalities in other countries such as Sweden, Norway, and Germany. Data for heights of Norwegian army recruits show impressive drops from 1761 to 1899–1903, 1922, and 1960–62.[20]

The Gini coefficients and lognormal slopes for the nine selected distributions of table 3.14 show little difference. One must generally conclude, given the strong possibility of measurement error, that relative inequality changed little. One of the statistical problems apparent in drawing generalizations stems from grouping errors. The first two distributions, derived from my sampling data, must be forced into the mold of the eight to sixteen published classes of the other distributions, a procedure I have performed, as explained in the source note of table 3.14. My preferred comparison is the Gini coefficient derived from sixteen classes shown in the next-to-last line of table 3.14, an arrangement obtained from an endeavor to achieve grouping uniformity. This row shows inequality for native-born in columns 2, 5, 7, and 9, of .0227, .0219, .0216, and .0217. There is slight evidence of decreasing inequality if one singles out these figures. This drop appears in more dramatic form when heights are sensitized, as shown in table 3.15. Yet, a glance at the plottings of the sensitized shares in figure 3.4 must lead one to conclude that any changes in equality of height were really quite minimal, certainly as contrasted to those I show for housing.

My presentation that highlights distribution of height as a normal curve certainly is not very novel. As long ago as 1863, E. B. Elliott plotted a frequency curve of the heights of 764 soldiers in the Army of the Potomac and, on the same chart, he plotted a normal curve fitted to his data. His measure of

18. I choose not to emphasize the fact that the sensitized Gini coefficients in table 3.15 for 1799–1819 are larger than those for the Civil War period. An element of taller men in my sample receives greater relative importance, obviously, in the sensitized version as compared to midpoints in frequency tables.

19. A computer run of data for 1799–1819 has $G(H)$ as .02246 for those of northern birth and .02364 for those of southern birth.

20. See Sweden (1969, pt. 1, population), for quinquenniel distributions from 1877 to 1949; Sandberg and Steckel (1980, 97) show a plot of the standard deviation of height from 1740 to 1880; for Norway, see Kill (1939, 64) and Udjus (1964, 47); for Germany, see Komlos (1990).

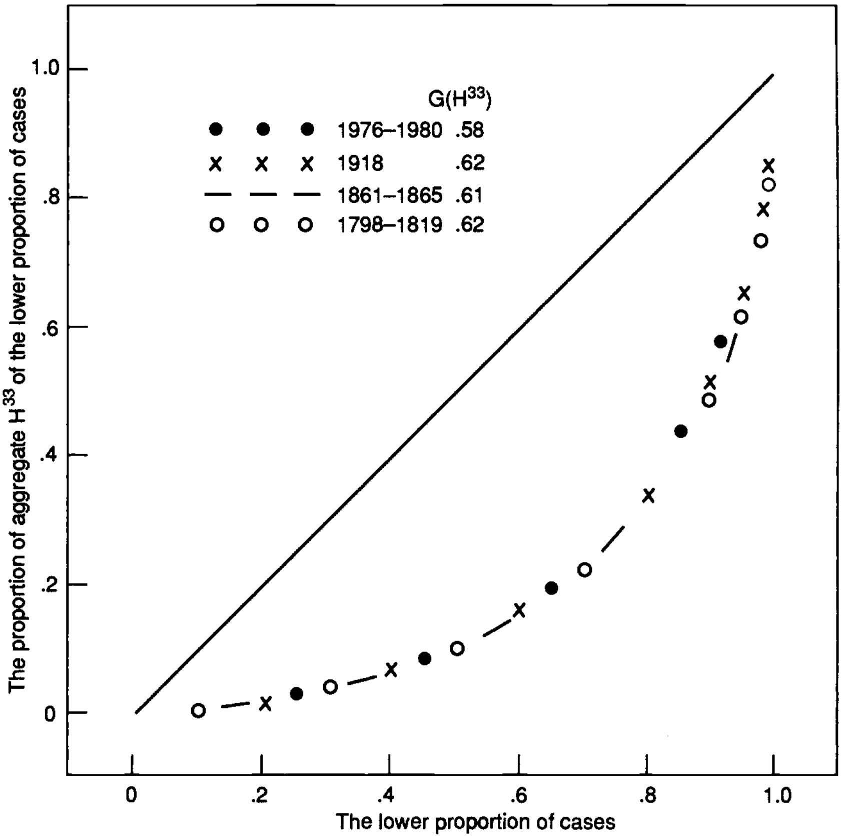

Fig. 3.4 Lorenz curves of the heights (H^{33}) of army recruits and draftees in three periods

Source: See table 3.15 for the four distributions numbered 2, 5, and 7 for native-born whites and 9 for the general population sample in 1976–80.

dispersion was the mean deviation, formulated at a time three decades before the development of the standard deviation. Since the mean deviation and standard deviation have an exact relationship for a normal curve, a reader can compute the coefficient of variation for his 764 cases from the parameters he reported for the 764. His coefficient of variation, .042, is for all practical purposes the same as mine derived from table 3.14 or table 3.15. Really, the only concept Elliott lacked was that of relative dispersion, as presented by Karl Pearson three decades later.[21]

Frankly, I had thought that relative inequality in heights in the United States

21. Elliott (1863, 40–44, diagram A). Pearson's work was cited by Yule (1919, 154, 160).

would have dropped because I suspected that income inequality has decreased since 1918. Perhaps changes in inequality in height, as distinguished from changes in mean height, are a result of a complex interaction of demographic, genetic, and environmental influences. In any case, one cannot readily offer height dispersion as an index of deteriorating income equality in the first half of the nineteenth century.

3.4.3 Literacy and Height

One glimpse of possible change in the characteristics of the short and the tall may be illustrated by using literacy rates. There is evidence that literacy improved more for the short than the tall, as illustrated with the following logistic equations for my sample of army recruits from 1799 to 1894: for 2,320 whites twenty-one and older with height (H) of 60 or more inches, and where ILL = 1 if illiterate, and ILL = 0 otherwise. 1799–1849:

$$\text{ILL} = 6.763 - \underset{(.0205)}{.1063}\, H;$$

$N = 1{,}658$, with 641 illiterate, and $P(\chi^2 = 28) < .001$; 1850–94:

$$\text{ILL} = -2.023 + \underset{(.0474)}{.0064}\, H;$$

$N = 662$, with 112 illiterate, and $P(\chi^2 = .02) < .892$. Among these, three occupational groups were distinguished, all showing similar drops in the force of illiteracy. Most impressive, of course, were all those with stated occupations other than farmer or laborer. There were similar findings for the subset of native-born recruits. We can say that at the beginning of the nineteenth century, literacy and height were strongly related. By the second half of the century, there was no advantage for the tall. The evidence is ever so slight that the short in some ways excelled in literacy among tradesmen and artisans if not among farmers or laborers.

3.4.4 The Poor in Farm Censuses

How desirable it would be to have distributions for food consumption, even if it were for adults only, that show inequalities in intakes of calories and nutritional quality for persons of different occupations and stature. Would there have been large variation in individual consumption? Would there have been a deprived group suffering from malnutrition? In the absence of consumption data, we might at least examine some production data at the farm level, particularly for the more marginal farms. The distribution of the number of cows possessed by farmers may be far removed from the inequality of milk consumption, but it does provide some gauge of the degree of sales necessary to achieve equality of intake.

South Carolina, 1850

Can anything be learned about the lower 40 percent from the censuses of agriculture for 1850–70? Do the farms of those with less than, say, twenty acres reveal characteristics of deprivation? Perhaps not. If these small farmers had holdings with little cash value, low crop production, and few cattle and livestock, they may be representative of individuals engaged in a multitude of other activities, including those measured by the value of home manufacturing, but overwhelmingly from income derived from alternative occupations.

I choose agricultural statistics for South Carolina as an example of what can be learned from farm data. In 1850 the state reported 29,967 farms and plantations, predominantly owned and managed by half of its 57,702 white males aged twenty and older. Farms arguably accounted for 52 percent of the white labor force, at best, and not 60 percent or more. Would a larger proportion of the white labor force have wished to possess acreage? Let us examine the lower tail of the acreage distribution for signs of deprivation. Marion Chandler and I drew a systematic sample of 861 farms from the state's agricultural census for investigation. Of these, 759 had improved acreage. The sample distribution of improved acreage plots neatly as a straight line on lognormal paper to ten acres at −1.1 standard deviations; at this point the line plummets rapidly (). The bottom 13 percent of farms, reporting almost no improved acreage, would have averaged about seven acres had the lognormal linearity continued among all cases.[22]

Yet the remarkable fact seems to be that there were so few farms below ten acres. Only 19 percent of farmers had less than twenty acres, and they appear to have been reasonably well equipped, if one considers some of their reported holdings, presented in table 3.16. It is possible that farmers reporting no acres had located on rental properties.

A lognormal plotting of the distribution of improved and unimproved acres shows a pattern similar to that just described for improved acres. This time the discontinuity appears at twenty-five acres, with 18 percent of farms below this level. Cash values were reported for only 733 of the 861 farms; the remaining 128 may have been rental, as distinguished from owner, properties, at least in part. The 733 exhibit an excellent lognormal linearity throughout the entire range. The 128 farms without stated cash value were more marginal, as shown in table 3.16, and appear to have experienced little activity in the economic sense. Yet almost all farmers owned a horse, a cow, and some livestock. Importantly, the 128 had an average of six acres in improved land. These, particularly, may have been occupied by persons with alternative employment. The distributions of cows and horses demonstrate lognormal linearity except for the 7–9 percent of farms reporting neither. The upshot is that the agricultural census yields few clues about income for all persons below the median.

22. Another extrapolation of the lognormal curve to about the level where all white adult males would have owned acreage would mean that the smallest farm would have had about one acre or a little less of improved land.

Table 3.16 Average Holdings of Marginal Farms in South Carolina in 1850

	Improved Acres		Cash Value of Farm	
	Under 20, or Not Stated	20 or More	$0, or Not Stated	$1 and Up
Number of farms	156	705	128	733
Improved acres	10.8	138	5.9	133
Unimproved acres	3.8	379	0	377
Cash value ($)	175	2,740	0	2,060
Cash value of farm implements and machinery ($)	32	131	15	130
Horses	1.2	3.3	1.3	3.2
Milch cows	2.3	6.4	1.9	6.3
Value of livestock ($)	135	494	124	482

Source: A systematic sample from the agricultural census, drawn and processed by Marion Chandler and Lee Soltow. At least part of those with no acreage or no cash value stated may have been rented properties.

Ohio Acreage in 1835

Comprehensive data sets for the production of marginal farmers for years before the census of 1850 are difficult to obtain. We should examine the acreage owned by Ohio farmers in 1835 for their implications concerning persons who were nonowners. We can compare counties with greater and smaller ownership rates and then make some inferences concerning marginal groups.

Stress has been placed on the great *regional* variation in landownership from county to county, township to township, and village to village. If the proportion owning land (LOP) was 80 percent in one township and only 20 percent in another not far away, wouldn't this indicate that land deprivation was not crucial? Otherwise, farm laborers would have migrated, lessening the dispersion in LOP. Some insight into the degree of ownership can be obtained from the set of 164,962 property values reported in Ohio in 1835.

One can obtain the number of property owners in each county by collating property values by the name of the owner. This procedure results in some difficulties because there can be several persons sharing a common name in a more densely populated county. It is impossible to conduct a genealogical study for all persons in Ohio; therefore it is necessary to accept some degree of measurement error in the collation of names. Separate estimates, based on collations at village level, township level, county level, a four-region level, and state level, all indicate general agreement in the patterns of inequality, no matter what level of collation is employed.[23] I choose to focus on the county-level arrangement.

The LOP in an Ohio county in 1835 is calculated by dividing the number of landowners in the county by an estimate of its population of adult males

23. This problem and others are discussed in Soltow (1987b).

twenty-one and older. This population is taken to be 25 percent of the geometric mean of the county's total population in 1830 and 1840. The resulting dispersion in this proportion is shown in table 3.17; three-eights of the state's population lived in counties where ownership rates were more than 60 percent. Land in these areas was spread very generously indeed among people. In about 5 percent of counties there was an abundance of property owners, and the vast majority of residents enjoyed ownership. In this paper, perhaps I should highlight the proportion without land if a major focus is to be on persons below median income. To consider those without land (or house ownership) as a deprived group, however, is somewhat tenuous, particularly for those who were artisans or tradesmen who rented.

What can be learned about areas with high ownership rates? One suspects that they would tend to be those counties well endowed, from the standpoint of overall wealth. Consider just one aspect of ongoing research in this area.

$$\text{LOP} = .495 + \underset{(.000123)}{.000256W};$$

$R^2 = .06$, $N = 73$ counties, where W is the aggregate value of real estate in the county, divided by its adult male population. Ownership rates did tend to be higher in wealthier counties. Those counties having a larger proportion of their population in agriculture (NAGRIC_T), as measured by the 1840 census, had wider spread in ownership but the relationship was rather weak.

$$\text{LOP} = .404 + .20\ \text{NAGRIC\_};$$

$R^2 = .03$, $N = 72$. This regression equation projects to limits of .40 for counties with no agricultural population and .60 for the entire population in agriculture. If Ohio followed this pattern over time and if it moved from 80 percent in agriculture to 70 percent in a decade, then it would experience a 2 percent drop in the LOP. This certainly does not mean that those just below median income need to have suffered from the standpoint of alternative employment. To a certain extent we can understand the meaning of "marginal" farms, shown by the case of South Carolina in 1850. One final perspective of Ohio in 1835 will be presented by partitioning counties into two groups, based on the LOP and by observing their lognormal configurations.

Table 3.18 distinguishes those forty-three "landless" counties in Ohio from the thirty with wide ownership, where "landless" in this context means the lower 40 percent who owned no land. Plots of the distributions of landed wealth among wealthholders for the two groups reveals no essential differences in slope, b (1.435 and 1.422), or relative inequality, G (.638 and .628), as measured by the Gini coefficient. It is almost as though the more marginal holders who were participating more broadly in the second group merely queued at the bottom of the parade of wealthholders, orderly blending in. It is somewhat tempting to think that if all persons had chosen to own or had been

Table 3.17 The Landowner Proportion (LOP) in the Seventy-three Counties in Ohio in 1835

LOP	Number of Counties		Proportion of Ohio's Population
.80 and up	5		.06
.60–.79	25		.31
.40–.59	36		.53
.29–.39	6		.09
.19 and less	1		.01
Total	73		1.00
Weighted population			
Mean		.56	
Gini coefficient		.14	
Median		.55	

Source: Ohio Tax records (1835); see also Soltow (1987b).

capable of owning land, the relative value distribution would have been lognormal, with a G of a little over .6.

Were persons in those counties with a lower ownership rate suffering in any sense relative to persons in counties with widespread ownership? At this date I am unable to interview those landless, but I can review some significant indicators, as listed at the bottom of table 3.18. In landless counties obviously there were fewer persons in the agricultural sector. They were older counties, as judged by formation dates, and tended to have more investment in manufacturing. Yet they did display more illiteracy and were to remain more illiterate. In these counties, school attendance was appreciably less. Nevertheless, overall average wealth among the rich and poor in succeeding decades was about the same for both groups. It is difficult to generalize about shares of lower groups on the basis of whether or not they owned land.

3.5 Household Clothing Production

Measurements have been made of the degree of inequality in family and individual condition with respect to saving (wealth), housing, and food and nutrition. Can anything be said about clothing? One can dream of a meter's being placed on a sample of persons on a Wednesday or a Sunday that would register the value of the clothing worn by each of them, in any year from 1798 to 1860. Alternatively, what would be the Gini coefficient for clothing expenditures for one year?

I must be content with an admittedly tangential presentation, one dealing with the variation in yards of cloth manufactured in individual households in New York State in 1825. This was a year when household production was very substantial relative to production in factories, more so than is stated in the censuses of 1835 and 1845. The distributions presented in table 3.19 do dem-

Table 3.18 Distribution of Real Estate (RE) among Owners Twenty-one and Older in Ohio in 1835, Classified by Counties with Low and High Landowning Proportions (LOP)

Real Estate Value ($)	43 Counties, Each with LOP <.60	30 Counties, Each with LOP ≥.60
50,000 and up	4	7
20,000–49,999	57	48
10,000–19,999	198	107
5,000–9,999	582	266
2,000–4,999	2,885	1,298
1,000–1,999	6,196	3,571
500–999	12,432	9,081
200–499	22,335	20,320
100–199	14,243	13,982
50–999	7,545	8,040
20–49	4,242	5,481
10–19	1,655	1,691
5–9	944	926
2–4	321	332
1	36	50
Total	73,675	65,110
RE ≥ 1		
Mean ($)	598	462
Inequality (*G*)	.638	.628
Lognormal chart		
R^2	.999	.994
Slope (*b*)	1.435	1.422
Number of nonowners	82,249	28,832
Males twenty-one and older	155,924	93,742
RE≥0		
Mean ($)	283	314
Inequality (*G*)	.829	.744
Lognormal chart		
R^2	.998	.990
Slope (*b*)	1.625	1.648
Weighted county averages		
Landowner proportion (LOP)	.472	.695
Proportion employed in agriculture, 1840	.694	.779
Proportion illiterate adults, 1840	.064	.040
School attendance per person aged 5–19, 1840	.284	.490
Daily newspapers per adult male, 1840	.039	.006
Proportion illiterate adult males, 1850	.056	.039
Capital investment in manufacturing, adult males, 1840 ($)	65	24
Cash value per acre of farms, 1850 ($)	28	20
Cash value per acre of farms, 1960 ($)	42	34
Total value of real estate, adult males, 1860 ($)	1,820	1,760

Table 3.18 (continued)

Real Estate Value ($)	43 Counties, Each with LOP <.60	30 Counties, Each with LOP ≥.60
Average latitude	39.8°	40.4°
Average longitude	82.8°	82.2°
Average year of county formation	1805	1809

Source: See table 3.17 and Inter-University Consortium for Political and Social Research (n.d.). The parameters for the lognormal chart were obtained from fitting the linear model to thirty-five and forty selected points for 1≤RE≤91,40 for RE≥1; for RE≥0, a sharp discontinuity appears for the ten-or-so points from RE of 1 to 90, and only those points on the chart above 90 were considered in the linear fitting.

Table 3.19 **Home-woven Yardage in New York in 1825: A sample of 1,775 Families and Households and 4,744 Adults in Thirteen Counties**

Total Yards per Household	Number of Households	Total Yards per Adult[a]	Number of Adults
1,000 and up	1		
500–999	2	500 and up	2
200–499	56	200–499	1
100–199	257	100–199	29
50–99	493	50–99	360
20–49	479	20–49	1,699
10–19	143	10–19	1,165
1–9	41	5–9	561
0	303	0.1–4	239
		0	688
Total	1,775		4,744

Inequality Coefficient

	Per Household	Per adult
G for yardage ≥ 0	.518	.487
G for yardage > 0	.418	.400

Source: New York (1825). Approximately every tenth page was sampled from extant records of this census, as available in the L.D.S. Genealogical Library, beginning with microfilm no. 806800 for Broome County.

[a]Adults were males eligible to vote and women sixteen and older. The distribution per adult assumes an equal distribution within each household.

A plotting of the thirteen counties (Broome, Chautauqua, Cortland, Herkimer, Jefferson, Lewis, Oneida, Orange, Steuben, Tioga, Tompkins, Washington, and Yates) among the fifty-five in the state shows a wide scatter of areas. The yardage per adult was 20.1 in the thirteen counties and 20.5 for the other forty-one counties, excluding New York, as reported in summary tables for 1825. Total yards is the simple addition of yards of fulled cloth, flannel, and other woolen cloth not fulled, and yards of linens, cotton, and other thin cloths manufactured "in the domestic way."

onstrate sizable inequality. They suggest that standards of living for clothing consumption may have varied significantly among households to the extent that they depended on this home production.

3.6 Alternative Measures of Inequality

In this paper I have singled out distribution measurements dealing only with wealth, housing expenditures or rents, literacy, schooling, number of children, stature, farm food production, and home-woven yardage. Nothing has been said about inequalities in medical treatment or life expectancy. Surely there must be alternative ways of measuring inequality; some of these can prove to be quite innovative and may yield annual distributions.

One could display distributions of the number of servants in households as a measure of well-being of the rich or possibly the degree of dependency of the poor. A tax on the value of carriages of various types owned by the select few percent of households indicates opulence in the early federal period. Distributions of horses owned indicates facility in transportation. Numbers of seats and differential pricing in theaters, and later on trains, certainly reflect class differences. Frequency tables of the number of watches and clocks convey notions of standards of living for various groups and how these standards changed as better methods of production and resulting lowering of prices were experienced. An extreme is a unique table classifying couples at marriage in five distinct economic classes in Amsterdam for each of some 250 years. It divulges the increased equality in the twentieth century, as opposed to inequalities existing from the seventeenth to the nineteenth centuries. Its *annual* distributions from 1829 to 1860 (and in some respects from 1818 to 1860) highlight rough constancy of relative inequality better than any statistical series I have developed (Soltow 1989b, 1989c).

Alternatively, landownership for persons classified by age indicates the rise in populations relative to land availability in such diverse areas as Connecticut and Norway over the last three or four centuries. Studies in the inequalities in the standard of living may be only beginning.

3.7 Conclusions

This paper has dealt with ten or so data sets depicting the distributions of wealth and income, or proxies for them such as educational level, marriage rates, and number of children, in households with and without wealth. The main emphasis has been on the degree of inequality, the disparity between rich and poor, appearing in these distributions for the United States in the nineteenth century. I also have tried to measure the change in inequality, but this was possible only in three facets: wealth, housing, and height.

Stated differently, the incomes of people have been viewed as the composition of saving (as measured by wealth), rent expenditure (as measured by dwelling values), food expenditure (as measured by height and household farm production), and clothing expenditure (as measured most indirectly by household cloth manufactured). An estimate of the income distribution, made from dwelling values in 1798, has a lognormal shape and a Gini coefficient of inequality (*G*) of about .6. This income inequality can be viewed as a synthesis of inequalities varying between a *G* of .8 or higher for saving (wealth) to one of .4 for food expenditure, and one can generate such a synthesis on the computer using lognormal assumptions (Soltow 1989a, 272–73).

An income inequality coefficient of .6 in 1798 means that there were large disparities at the time, substantially larger than today, when the coefficient is rising, but still is less than .5 (U.S. Bureau of the Census 1990, 30; Maxwell 1990). Is this dwelling-derived estimate of income inequality in 1798 too large? I believe the estimate is reasonable, considering the fact that the relative distribution of the sizes of families was substantial at the time, with a *G* of about .3. There was a strong degree of variation in the average income of counties and smaller area units. Adjustment for this variation decreases *G* from .60 to .55. Consider also the fact that the economy at that time was dominated by seasonal movement in a rural setting subject to floods, drought, insects, fires, sickness, and the remainder of the disruptions due to war. There were great differences attributable to farm size, land fertility, and land terrain. The effects of inheritance on wealth, dependent on past accumulation, were large. Traces of the influence of primogeniture persisted in several states. And slavery strengthened inequality of income among the free.

Did the relative inequality of income change before the Civil War? Probably not much, if at all, but conclusions must be derived from several different sources with varying degrees of precision. Consider first the distribution of wealth, our most complete time series. Figures 3.1 and 3.2 demonstrate that *G*'s for inequality of wealth in real estate were .6 to .65 from 1798 to 1870 for those possessing estates. Coefficients for real and personal estate in 1860 and 1870, and probably earlier, were roughly the same. It is true that the proportion of adult males owning land decreased from .5 in 1798 to below .45 two generations later.

Wealth distributions are generally silent about the 40 percent of individuals with no wealth. We do know from complete census enumerations that those without wealth were less likely to be married, less likely to have children, and that they had lower literacy rates and children with less schooling. Yet these statistics concerning the poor do not seem particularly disturbing to me. The shanty counts for Ontario in 1851 do impress one with the fact that there were large minority groups who lived in very humble circumstances, at least compared to most lower groups at the present time.

It is from housing data for New York State that we obtain our most authoritative facts about the relative shares of households. The lowest 40 percent of homes (excluding New York and Kings counties) accounted for 7 percent of aggregate housing value in 1798, 8 percent of the value in 1875, and 18 percent of the value in 1980. It is these important quantifications that lead me to conclude that there was little change in inequality in the period of our interest, and that the poor really had very little compared to the middle and top groups in the nineteenth century.

Finally, we must not ignore distributions of heights, in part because they are available for *all* recruits in roughly the years 1812, 1864, and 1918 and for a sample of all adult males for 1976–80. I find that relative inequality in these figures, the Lorenz curves in these years, approximately duplicate each other. There is no evidence of deterioration among lower groups relative to upper groups to the extent that nutrition and food consumption do affect height distributions. The New York dwelling and the height measures both signify little or no inequality change in the nineteenth century. The two sources are inconsistent as they signal inequality change for the twentieth century. Probably food consumption was much less sensitive than was housing to the relative rise in income of lower groups.

There are many ways of measuring inequalities in the standard of living in any one year. Some of these are suggested in this paper as possible areas to be investigated in the future. Surely additional and alternative measurements must be devised. The study of inequalities in the standard of living is only in its infancy. The relative shares available to the rich and the poor in the nineteenth century must be a fundamental part of the documentation measuring economic growth.

References

Aitchison, J., and J. A. C. Brown. 1969. *The lognormal distribution*. Cambridge: Cambridge University Press.

Amsterdam. 1923. *Statistiek der Bevolking van Amsterdam tot 1921*. No. 67. Amsterdam: het Bureau van Statistiek der Gemeente Amsterdam.

Antwerp. 1799. Stadsarchief: Matrice de role d'Anvers, fonciers, ville. Sections 1–4.

Bailyn, B. 1986. *Voyagers to the West*. New York: Alfred A. Knopf.

Baxter, J. 1875. *Statistics, medical and anthropological, of the Provost Marshall General's bureau, derived from records of the examination for military service in the armies of the United States during the late War of the Rebellion of over a million recruits, drafted men, substitutes, and enrolled men*. Vols. 1 and 2. Washington, D.C.: U.S. War Department.

Canada. 1851–52. *Appendix to census of Canada, Upper Canada*. National Archives of Canada. Microfilm C-11, 712, to C-11, 762.

Davenport, C. B., and A. G. Love. 1921. *Statistics: Army anthropology*. Vol. 15, part 1. Washington, D.C.: GPO.

David, P. A., and P. Solar. 1977. A bicentenary contribution to the history of the cost of living in America. *Research in Economic History* 2: 1–80.

De Belder, J. 1977. De behuizing te Antwerpen op het einde van de XVIIIde eeuw. *Belgisch tijdschrift voor niewste geschiedenis* 8: 367–446.

Elliot, E. B. 1863. *On the military statistics of the United States of America*. Berlin: Decker.

Eveleth, P. B., and J. M. Tanner. 1976. *Worldwide variation in human growth*. Cambridge: Cambridge University Press.

Fogel, R. W. 1986. Nutrition and the decline in mortality since 1700: Some preliminary findings. In *Long-term factors in American economic growth*, ed. S. L. Engerman and R. E. Gallman. NBER Studies in Income and Wealth, 51. Chicago: University of Chicago Press.

Gould, B. A. 1979. *Investigations in the military and anthropological statistics of American Soldiers*. New York: Arno Press. Reprint of U.S. Sanitary Commission's memoirs of the War of the Rebellion, vol. 2. New York: Hurd and Houghton, 1869.

Great Britain. 1805. *Parliamentary papers, 1803–1804*. House of Commons, vol. 13.

Humphreys, N. 1887. Class mortality statistics. *Journal of the Royal Statistical Society* 50: 255–85.

Inter-University Consortium for Political and Social Research. N.d. *Historical, demographic, economic, and social data: The United States, 1790–1970*. Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research.

Kearl, J. R., and C. L. Pope. 1986. Choices, rents, and luck: Economic mobility of nineteenth-century Utah households. In *Long-term factors in American economic growth*, ed. S. L. Engerman and R. E. Gallman. NBER Studies in Income and Wealth, 51. Chicago: University of Chicago Press.

Kidwell, C., and M. Christman. 1974. *Suiting everyone: The democratization of clothing in America*. Washington, D.C.: Smithsonian Institution Press.

Kill, V. 1939. *Stature and growth of Norwegian men during the past two hundred years*. Oslo: hos J. Dybwad.

Komlos, J. 1990. Height and social status in eighteenth-century Germany. *Journal of Interdisciplinary History* 20: 607–721.

Lievense, E. N.d. Gronnen en methoden voor het onderzoek in archiven naar de geschiedenis van huizen binnen de singel gracht. Gemeentearchief Amsterdam. GAA call no. SZ.

Lis, C. 1986. *Social change and the labouring poor: Antwerp, 1770–1860*. New Haven: Yale University Press.

Margo, R. A., and R. H. Steckel. 1983. Heights of native-born whites during the antebellum period. *Journal of Economic History* 43: 167–74.

Martin, E. W. 1942. *The standard of living in 1860*. Chicago: University of Chicago Press.

Maxwell, N. L. 1990. *Income inequality in the United States, 1947–1985*. Westport, Conn.: Greenwood Press.

New York. 1825. *State census*. L.D.S. Genealogical Library, Salt Lake City. Microfilm.

New York. Secretary of State. 1855. *Census of the State of New York for 1855*. Albany: C. Van Benthuysen, 1857.

———. Superintendent of the Census. 1875. *Census of the State of New York for 1875*. Albany: Weed, Parsons, and Company, Printers, 1877.

Ogle, W. 1887. Summary of several male life tables. *Journal of the Royal Statistical Society* 50: 648–52.

Ohio. 1835. Ohio Historical Society Archives. Data recorded from tax duplicates by G. Pettey. Ohio State Historical Library, Columbus.

Pope, C. L. 1989. Households on the American frontier: The distribution of income and wealth in Utah, 1850–1900. In *Markets in history*, ed. D. Galenson. Cambridge: Cambridge University Press.

Reyniers-Defourny, R. 1979. *De bevolkingstelling van Antwerp in de France*. Stadsarchief, Antwerp. Mimeo.

Sandberg, L. G., and R. H. Steckel. 1980. Soldier, soldier, what made you grow so tall? *Economy and History* 23: 91–105.

Sen, A. 1976. Poverty: An ordinal approach to measurement. *Econometrica* 44: 219–31.
Sheppard, W. F. 1897. On the calculation of the average square, cube, etc., of a large number of magnitudes. *Journal of the Royal Statistical Society* 60: 698–703.
Soltow, L. 1971. An index of the rich and poor in Scotland. *Scottish Journal of Political Economy* 18: 49–67.
———. 1975. *Men and wealth in the United States, 1850–1870.* New Haven: Yale University Press.
———. 1981. Horse owners in Kentucky in 1800. *Register of the Kentucky Historical Society* 79: 203–10.
———. 1982. Male inheritance expectations in the United States in 1870. *Review of Economics and Statistics* 64: 252–60.
———. 1984. Wealth inequality in the United States in 1798 and 1860. *Review of Economics and Statistics* 66: 444–51.
———. 1987a. Distribution of income in the United States in 1798: Estimates based on the federal housing inventory. *Review of Economics and Statistics* 69: 181–85.
———. 1987b. Tocqueville's view of the Northwest in 1835: Ohio a generation after settlement. In *Essays on the economy of the old Northwest,* ed. D. C. Klingaman and R. K. Vedder. Athens: Ohio University Press.
———. 1989a. *Distribution of wealth and income in the United States in 1798.* Pittsburgh: University of Pittsburgh Press.
———. 1989b. Income and wealth inequality in Amsterdam, 1585–1805. *Economish-en Sociall-historisch Jaarboek* 52: 72–95.
———. 1989c. The rich and destitute in Sweden, 1805–1855: A test of Tocqueville's inequality hypotheses. *Economic History Review,* 2d ser., 42: 43–63.
———. 1990. Inequality of wealth in land in Scotland in the eighteenth century. *Scottish Economic and Social History* 10: 38–60.
Soltow, L., and D. L. May. 1979. The distribution of Mormon wealth and income in 1857. *Explorations in Economic History* 16: 151–62.
Soltow, L., and E. Stevens. 1981. *The rise of literacy and the common school in the United States.* Chicago: University of Chicago Press.
Steckel, R. H. 1983. Height and per capita income. *Historical Methods* 16: 1–70.
Sweden. 1969. *Historical statistics of Sweden.* Part 1, *Population.* Stockholm: Statistical Centralbureau.
Tocqueville, A. de. 1966. *Democracy in America.* Henry Reeve text. New York: Alfred A. Knopf.
Udjus, G. U. 1964. *Anthropometrical changes in Norwegian men in the twentieth century.* Oslo: Universitetsforlaget.
United States. 1987. *Vital and health statistics.* U.S. National Center for Health Statistics, ser. 11, no. 238. Washington, D.C.: GPO, October.
U.S. Bureau of the Census. 1984. *1980 Census of Housing.* Washington, D.C.: GPO.
———. 1990. Money, income, and poverty status in the United States, 1989. In *Current population reports,* ser. P-60, no. 168. Washington, D.C.: GPO, 1991.
Van Wieringen, J. C. 1979. Secular growth changes and environment: An analysis of developments in the Netherlands, 1850–1978. *Collegium Antropologicum* 31: 35–48.
———. 1986. Secular growth changes. In *Human Growth,* 2d edition, ed. F. Falkner and J. M. Tanner, vol. 3. New York: Plenum Press.
Yule, G. U. 1919. *An introduction to the theory of statistics.* London: Charles Griffin and Co.

Comment Clayne L. Pope

This paper adds to the long list of Lee Soltow's contributions to our knowledge concerning the changing distribution of economic rewards across time and space. Had I (and I suspect there are many others like me) paid Soltow a royalty for every occasion when I have opened *Men and Wealth in the United States* to make a comparison or check a number, he would be in the rich tail of the wealth distribution himself.[1] His contributions to the study of the trends in the distributions of income and wealth, use of the federal census manuscripts, and international comparisons of inequality are well known. He has brought new and useful sources of evidence on inequality, most recently the housing inventory of 1798, to the attention of others.

In this paper he has set a challenging task for himself and his readers—better understanding of the economic conditions or standard of living of the poor. Most of the papers of this volume are directed toward changes in the mean level of living. The focus of this paper is the poorest 40 percent of the population. While he does use the familiar census manuscripts with their estimates of household wealth and real estate, Soltow's search for evidence on this issue leads him from the shanties of Ontario and the tenements of New York, to the marriage arrangements of Amsterdam, and on to the correlation of literacy and heights of military recruits. I interpret this energetic search in disparate sources as evidence of the difficulty in gaining substantive knowledge on either the level of living of the poor or their share of the aggregate consumption or income. In spite of Soltow's clear expertise and his diligence, this paper reinforces the impression that we have a long way to go before we can speak with confidence about the standard of living of the poor or their share of the economic pie. Our most commonly accessible measures of economic status in the past are occupation and wealth. The census and other sources often give occupation while probate inventories, tax rolls, and some census enumerations give wealth. Unfortunately, neither occupation nor wealth tell us much about the relative position of the poor. The very poor have no measured wealth, and they share the occupation of "laborer" with many whose economic status is considerably higher. Soltow's search for the poor through unconventional sources is understandable though not necessarily fully successful.

One methodological point might be useful before consideration of the substance of the paper. Soltow compares the various distributions of economic success to the lognormal distribution. The lognormal is second only to the

Clayne L. Pope is professor of economics at Brigham Young University and a research associate of the National Bureau of Economic Research.

1. Lee Soltow, *Men and Wealth in the United States, 1850–1870* (New Haven: Yale University Press, 1975).

normal distribution in its usefulness.[2] If it may be assumed that a distribution is lognormal, counterfactual statements may be more easily examined because one has a simple two-parameter statistical model for the distribution.

It is possible to test statistically how well a lognormal distribution fits a body of data. The lognormal, a two-parameter distribution, is a special case of a three-parameter generalized gamma distribution which is, in turn, a special case of a four-parameter generalized beta distribution.[3] One could estimate the more general three- or four-parameter distribution and compare the estimated parameters to the restrictions that reduce the more general distributional specification to the two-parameter lognormal distribution.[4] Statistical tests may then be used to compare the nested statistical models to see if the use of a lognormal distribution sacrifices too much precision.[5]

For many issues, the fit of the data to any model distribution is not relevant. That is, we are simply interested in the actual size distribution, the share of the poorest fifth of the population, the computed Gini coefficient, and so on. The model distributions are useful for counterfactual statements, adjustments to the data or estimation of covariates of the distribution. For many of the issues in Soltow's paper, the goodness of fit of the data to the lognormal specification is not particularly germane because we are simply interested in the distribution per se.

The most serious difficulty with application of the lognormal distribution to wealth, or to a much lesser extent income, is the fact that many households report no wealth or income. Wealth distributions always have considerable mass at zero, and the fits to the lognormal distribution exclude zero values, as Soltow is careful to tell us. The zero wealth values present a serious problem

2. For discussions of the distribution and its applications see Edwin L. Crow and Kunio Shimizu, *Lognormal Distributions: Theory and Applications* (New York: Marcel Dekker, 1988), and J. Aitchison and J. A. C. Brown, *the Lognormal Distribution* (Cambridge: Cambridge University press, 1966).

3. James McDonald, "Some Generalized Functions for the Size Distribution of Income," *Econometrica* 52, no. 3 (May 1984): 647–63.

4. For example, the generalized gamma

$$f(y; a, \beta, p) = \frac{ay^{ap-1}e^{-(y/\beta)^a}}{\beta^{ap}\Gamma(p)}$$

becomes lognormal if

$$\beta^a = \sigma^2 a^2 \quad \text{and} \quad p = \frac{a\mu + 1}{\beta^a}.$$

5. When my colleague James McDonald applied this approach to grouped data for family income in the United States in 1970, 1975, and 1980, he found some differences between the performance of the four- and three-parameter specifications, with the generalized beta (four-parameter) doing significantly better. He also found that distributions such as the Weibull and the Singh-Maddala fit better than the lognormal. The error in the Gini coefficient implied by use of the lognormal ranged from 8 percent to 17 percent. See McDonald, "Some Generalized Functions." However, the lognormal is convenient. For historical issues, data problems are clearly more pressing than the differences implicit in distributional assumptions.

in most research on the wealthholding of households. The zero values are a mixture of households that truly have no wealth and households with little wealth. Inclusion or exclusion of zero values usually changes the coefficients of regressions explaining wealth and always changes the estimates of the level of inequality. This problem, which I view as a heaping problem, does not have a satisfactory solution as yet. Model distributions may be of some help in solving the problem of reported zeros in the data.

Each of the measures of economic success reviewed in this paper approaches the issue of poverty from a different vantage point. The parable of the blind men and the elephant serves as a warning. Each source may be producing information about a different aspect of poverty. In this discussion, I would like to pose questions or difficulties in the use of three of the data sources—height, housing values, and wealth.

Height

The distribution of heights is the result of a mixture of differences within the population in genetics and nutritional status. Note that the distribution of heights does not correspond closely to a distribution of economic status. That is, height of 77 inches compared to 70 inches is not analogous to an income of $50,000 compared to one of $35,000. To quote Tom Thumb from the musical *Barnum,* "Bigger isn't Better." The link of height and poverty works through the gap between height determined solely by genetic potential and actual height. Some of the individuals in the short tail of the height distribution are close to genetic potential and have not suffered any particular poverty. Others are short because of poor nutritional status that may well be the result of poverty. The point is that height is not a very good predictor of poverty except in the extreme. Knowledge about height inequality does not translate easily into knowledge about economic inequality.

Many height distributions have been compiled from censored or truncated data sets. Most researchers have used the assumption that heights are normally distributed to recover the actual distribution from the censored data. It is hard to see how that procedure would lead to a normal distribution with more dispersion if the right-hand side of the distribution is largely determined by genetics. It may be that height distributions, useful for trends in the mean standard of living, will not be directly useful for the study of the very poor until we have much greater precision in the measurement of the height distribution and its deviation from normality.

The more promising use of heights is likely to continue to be multivariate analysis that is used to identify groups with varying proportions of poverty—urban, rural, geographical regions, or occupational groups. Along this line, Soltow's results on height and literacy are intriguing. However, the nineteenth-century cycle in mean height may be playing a role in the regressions.

Housing

The distribution of housing values represented by the distributions for New York in 1798, 1875, and 1980 are related to an important element of poverty. Certainly, we would be willing to use housing today as a useful indicator of poverty. But there are substantial adjustments to be made in moving from the value of housing to the consumption of housing services. The problem is illustrated in table 3.6. There are virtually no dwellings in New York or Kings counties in the lowest two classes in 1798 or in the lowest three classes in 1875. I would not be ready to conclude that housing services were better in those counties. We know that locational rent in large cities can be substantial. Location certainly confers advantages that raise wages. Unfortunately, the poor are more likely to be unemployed or infirm and unable to capitalize on locational advantages that make their wages and earnings higher to offset the higher cost of their housing services. Thus, the lower level of inequality of housing values in New York and Kings counties compared to the other counties may be illusory.

Wealth

In one sense, wealth is not particularly useful for analysis of poverty because the poor are unlikely to have recorded wealth. In 1860, Soltow found that 38 percent of households owned no wealth.[6] If we could assume that all who have no wealth were poor, the wealth distributions would serve well, but such an assumption is unwarranted. Wealth-income ratios vary systematically over the life cycle. Indeed, part of the life cycle involves the transfer of human capital into nonhuman capital for bequest or later consumption. This omission of human capital from wealth estimates limits the usefulness of wealth distributions for the study of poverty. Even beyond the life cycle, there is also a great deal of unexplained variation in the wealth-income ratio. The connection between wealth and income is probably least firm in urban areas where renting of housing is more common. Urban households with no wealth may well include many households living comfortably beyond poverty.

Wealth distributions are useful, however, for the study of mobility of households out of poverty or at least low wealthholding. If households are linked over several different years of observation, the group that remain with little or no wealth can be identified. For example, there was substantial mobility in nineteenth century Utah with very few of the poorest households in one census year still poor in the succeeding census.[7] We need more evidence on the extent of occupational and wealth mobility for poor households. Unfortunately, the data sets useful for the measurement of mobility are rare and difficult to create.

6. Soltow, *Men and Wealth*, 60.

7. See J. R. Kearl and Clayne L. Pope, "Wealth Mobility: The Missing Element," *Journal of Interdisciplinary History* 8 (Winter 1983): 461–88.

Soltow's paper makes clear how difficult it is to find historical information on the poorest segment of the population. Sources such as relief rolls or orphanage records may yield more information, but we are not likely ever to have much quantitative data on the condition of the very poor. However, the problem may not be as serious as first appearances suggest. Throughout this paper, Soltow reminds us of how little change in inequality has occurred over the past two centuries (with the exception of dwelling values). With rather stable levels of inequality, changes in the level of living for the poor must correspond quite closely to changes in the mean standard of living for the full population. In terms of height and nutrition, it would seem reasonable that the status of the poor increased at the same or a somewhat faster rate than the rate of change of the general mean. Consequently, our attention can, for the most part, be on the trend in the average level because the distribution was changing very slowly if at all.

Once the movement of the mean level of living is established, attention could be shifted to the mobility of the households out of poverty. Our view of poverty will clearly be conditioned on the turnover of the poor or the extent of movement out of the poorest groups of the economy. I applaud Soltow's energy in the collection and analysis of various distributions of economic status. But I believe most of our research should be directed toward better measurement of the average level of living and the mobility of households within the generally stable distributions of economic status.

This Page Intentionally Left Blank