

ON THE PERSISTENT STRUCTURE OF METROPOLITAN INCOME INEQUALITY, 1900-1980

*Joseph Persky and Mo-Yin Tam**

Abstract—A unique data set for 1900 is used to construct estimates of mean incomes and Gini coefficients for "normal families" in "principal cities" of 33 states. Southern states showed much higher inequality levels than northern ones. *Ceteris paribus*, income levels were inversely related to inequality, while non-white population shares were positively related to inequality. Inequality rose for the very largest areas, but somewhat surprisingly fell with higher illiteracy rates. The Census PUS for 1980 was used to construct state samples based as closely as possible on the definitions of "normal families" and "principal cities" used in 1900. The general cross-sectional pattern of inequality in 1980 was remarkably similar to that of 1900, although there seems to have been a convergence of inequality levels and a declining significance of race and region.

I. INTRODUCTION

For years now, economists and regional scientists have used state or metropolitan area income distributions to research the determinants of intra-regional income inequality (Al-Samarrie and Miller 1967; Sale 1974; Danziger 1976; Garofalo and Fogarty 1979; Nord 1980, 1990; Amos 1988). The variation of income inequality across states and metropolitan areas allows a natural laboratory for exploring the factors that promote a greater equality of family income. This work has largely focussed on the post-World War II period, since little income distribution data is available before 1940. The purpose of this paper is to use a unique sample of family incomes in 1900 to carry out a cross-sectional analysis of inequality in that year. We attempt to provide a historical anchor to the cross-sectional approach to the determinants of income inequality.

The keystone of our approach is a set of data for 1900 reported in *The Eighteenth Annual Report of the Commissioner of Labor: Cost of Living and Retail Prices of Food* (1903). Both Lebergott (1984, 498) and before him Kravis (1962, 209-210) have used data from this report in their analysis of national income inequality. Neither researcher attempted a regional analysis. Yet surprisingly, the *Report* presented, in addition to national income distributions, distributions for families in the principal cities of 33 states. These data make possible our cross-sectional approach to income inequality in 1900.

*Department of Economics, University of Illinois at Chicago.

As with many historical studies, our primary interest in the past is to compare and contrast it with the present. In this exploratory effort, we consider several of the most important correlates of inequality identified in the cross-sectional literature. In particular, we can ask how the relation of metropolitan income inequality to such variables as average income, racial composition, and industrial structure has changed over the century. Have the correlates of inequality remained largely the same over the years, or can we observe structural changes? After accounting for important correlates, does a significant regional effect remain unexplained and if so, how has that effect changed?

The next section takes up methodological issues. Most importantly, it considers how to make meaningful comparisons between 1900 and more recent years. In Section III the basic geography of inequality then and now is described. Section IV explores the relation between average income levels and family income inequality in the two years. This allows us to comment on Kuznets's famous hypothesis that inequality is likely to first rise and then fall with the level of development. Section V considers the role of race, education, and industrial structure, while the last section gives our conclusions and suggests directions for future work.

II. METHODOLOGY

When we first came across Stanley Lebergott's (1976) estimates of the U.S. income distribution for 1900, we were struck by his assertion that data from the *Report* were better than any subsequent year down to 1935-36. This is certainly the case, but the 1900 data cannot be directly compared to those from the post-World War II period. Unfortunately, the families included in the income distributions of the *Report* are far from a random sample. First, they are drawn from "principal industrial centers." Second, the data only include families with heads "employed as wage workers and at small salaries" (*Report* 1903, 15). Finally, the income distributions are reported only for "normal" families, those which had

1. The husband at work.
2. A wife.
3. Not more than five children, and none more than 14 years of age.
4. No dependent boarder, lodger, or servant.
5. Expenditures for rent, fuel, lighting, food, clothing, and sundries (*Report* 1903, 18).¹

The *Report* surveyed 11,156 "normal" families in the country.

Researchers using the *Report* have followed two very different strategies to insure comparability to more recent data. One can either attempt to supplement the *Report* with other sources to estimate an income distribution representative of a broader population, or one can pare down a recent survey to match the characteristics of the earlier sample of "normal" families.

Lebergott (1976) took the income distribution of these "normal" families and combined it with distributions estimated for farm families, the non-farm self-employed, and for families headed by women. He worked these up into an income distribution for the nation as a whole. This exercise requires numerous assumptions about the incomes of the most prosperous and the most destitute families. Whatever the adequacy of Lebergott's efforts, the data necessary to make such adjustments at the state level simply do not exist.

If the mountain won't come to Muhammad, Muhammad must go to the mountain. If we cannot adjust our income distributions from 1900 to current definitions, perhaps we can adjust more recent data to match that of the *Report* for 1900. The population of "normal" families in "principal industrial centers," although certainly not representative of all families, still is of considerable interest in and of itself. At the national level, this approach was taken by Irving Kravis (1962), who attempted to construct a sample from the 1950 BLS Survey comparable to the sample of "normal" families for 1900.

Kravis was primarily interested in commenting on time trends in the level of national inequality. His procedure has been criticized by Williamson and Lindert (1980, 91), who questioned whether Kravis really achieved a comparable sample.² Kravis was in fact somewhat limited by the structure of the data available to him, although we think he did a reasonable job. Using data from the 1980 Public Use Sample, we have somewhat greater freedom in defining the precise characteristics of the sample to be used. As discussed below, we have devoted considerable effort to achieving a solid match. However, even if we have been less than fully successful, our project is likely to be far less sensitive than Kravis's to such sources of error. A modest change in the definition of a sample can make a considerable difference in the level of inequality measured for that sample and hence in the change in inequality measured over time. Such sampling problems are much less likely to affect any cross-sectional patterns of inequality. Our primary interest here is to explore the persistence, or lack thereof, in the determinants of metropolitan income inequality. At worst we expect any mismatching of samples to influence only the magnitude of measured effects and not their statistical significance.

The abundant micro-data available from the Census of 1980 allows almost unlimited freedom for drawing out subsamples and constructing their income dis-

tributions. Our strategy becomes one of attempting to reproduce for 1980 state family income distributions comparable to those in the *Report* for 1900.³

The criteria we used in constructing the distributions for 1980 were that an included family must have both a husband and a wife, the husband must have been employed at the time of the Census, and the family had to live in an urbanized area (as defined by the 1980 Census).⁴ The 1900 data excluded families in which the husband worked for a salary greater than \$1,200, although no limit applied to wage workers. The tabulations available allow us to also exclude the wage workers earning more than \$1,200 in order to get a uniform income cutoff. The question became one of choosing a corresponding income cutoff for 1980. While somewhat arbitrary, we chose to multiply \$1,200 by the ratio of average U.S. family income in 1980 (about \$24,000) to average family income in 1900 (about \$900).⁵ This gave a figure of about \$32,000 which we rounded down to \$30,000.

We impose no other constraints on the 1980 families. Family size and the presence of boarders seem far less important now than in 1900. A bit more questionably we did not insist that families be renters in 1980. This also seemed of secondary significance, given the huge increases in home ownership since 1900. Finally, we did not exclude families with children older than 14.

The characteristics of the sample in 1980 and of the original set of "normal" families should make it clear that the inequality measures determined from these data for either year and for any state cannot be taken as an estimate of inequality for the state's metropolitan population as a whole. Even if the match here was perfect, we could not infer from changes in the inequality measured for these samples the overall change in family inequality. Most obviously, the construction of these samples rules out important changes in family inequality traceable to changes in the prevalence of single-parent families.

Our hope here is to use geographic variation in inequality measures across states to isolate for each year, 1900 and 1980, important determinants of inequality. Of course, the nature of the two samples puts constraints on this project. For example, we cannot very well explore the importance of single-parent status if single-parent families are systematically excluded from the two samples. Thus, we must accept from the beginning that we can explore only a subset of the important determinants of family income inequality.

The 1980 distributions were derived from the 1-in-100 C-Sample of Public-Use Microdata Sample (Census 1983). Unfortunately, all states are not identified in these data. The unidentified states are gathered into subregional state groups. On the other hand, the *Report* identified only 33 states for 1900. Putting these together as best we can to match the data for 1980 leaves 25 observations, 21

states, and four multistate groups for which both 1900 and 1980 income distributions could be estimated.⁶

The Gini coefficients for each state (or multistate group) were estimated from simple linear interpolations over \$100 intervals for 1900 and \$5,000 intervals for 1980. Because of the availability of micro-data, we could have had a much more precise distribution for 1980, but we wished to keep the estimation technique comparable to the earlier year.

III. THE GEOGRAPHY OF "NORMAL" FAMILY INCOME DISTRIBUTIONS

In 1900, the mean income for all "normal" families in our sample was \$647. The Gini coefficient for this same group of families was 15.9.⁷ These data clearly are not representative of the overall distribution of income in the nation. Lebergott reports an average family income of \$899, and we estimate, on the basis of Lebergott's data, a Gini coefficient of 39.0 for the entire nation in that year. Some of this difference undoubtedly derives from substantial interstate differentials in average family income. A more important portion represents the effect of income truncation in the distribution used here and the exclusion of farm families, the self-employed, and families headed by women.

"Normal" families in 1980 obviously were much richer than their counterparts in 1900, with a mean income of \$19,444. But, they also show a more unequal income distribution, with a Gini coefficient of 20.4. These can be compared to the Census Bureau's figures for the total family population in 1980. They report a mean of \$23,974 and a Gini of 36.5. Thus, while inequality in our group of "normal" families seems to have risen, inequality among all families probably fell.

As discussed above, we are not primarily concerned with estimating the change in inequality levels over time. Nevertheless, the data are suggestive. Our results for the constructed sample in 1980 can be compared to Kravis's estimates for his own constructed sample⁸ based on the 1950 Bureau of Labor Statistics survey. He reports a Gini coefficient of 19. The Current Population Survey figure for all families for 1950 was 37.9. These figures are consistent with the tendencies suggested above. On the other hand, we should note that there are strong indications that the 1900 *Report* heavily oversampled families with heads in manufacturing.⁹ Since manufacturing industries have traditionally presented more egalitarian income distributions than most other activities, the bias in the *Report* may well be downward.

Focussing now on the 1900 state data for "normal" families in large urban areas, we find an average (across states) mean income of \$626 and an average Gini coefficient of 16.7 (see Table 1). There was considerable variation in the degree of inequality for "normal" families across states. The coefficient of variation of the state Gini coefficients was 0.25.

The most obvious pattern in the state Gini coefficients is geographic. There was a large difference between the measured inequality for "normal" families in the principal cities in the states of the South and those in the rest of the country. The income distributions for "normal" families in the South showed considerably greater inequality than in the former, reaching up to 32.5 for South Carolina. Colorado showed the lowest Gini coefficient, 12.1.

For our 1980 sample, the average mean income across states was \$19,454. The average Gini coefficient for our states was 20.2 (again, see Table 1). Unlike 1900, the dispersion of the 1980 Gini coefficients was quite narrow. The coefficient of variation across all states was only 0.08. The range in 1980 ran from a minimum of 17.6 for Wisconsin to a maximum of 23.4 for Texas. Thus, while a representative family in 1980 would perceive more inequality among "normal" families in its own state than such a family in 1900, its own state's income distribution would be more or less indistinguishable from those of other states.¹⁰

If states had more similar income distributions in 1980 than in 1900, there is still a marked persistence of the geographic pattern of inequality across the years. The rank correlation of state Gini coefficients in 1900 with those of 1980 is 0.55, which largely reflects the persistence of higher inequality among "normal" families in southern states as compared to northern ones. Although muted, the pattern of higher southern inequality remained surprisingly consistent. On the other hand, western states that scored well in 1900 tended to move down by 1980. The largest shift in position occurred in California, which ranked seventh (with first being most equal) in 1900 and nineteenth in 1980. Colorado also fell precipitously from first to thirteenth, while Washington fell from second to sixth. These were the only western states included in the data, although Texas, which following the Census we label as southern, fell from fifteenth to twenty-fifth, the very bottom.

IV. LEVEL OF DEVELOPMENT AND METROPOLITAN INCOME INEQUALITY

In 1900, southern states had Gini coefficients that averaged 5.7 points, or about 40 percent more than the coefficients for northern states. This difference is the most striking characteristic of the early data. A similar but smaller difference appears in the 1980 coefficients—2.2 points, or 12 percent of the northern average.

TABLE 1
Mean "Normal Family" Incomes and Ginis by State, 1900 and 1980

State	1900 Mean	1900 Gini	1980 Mean	1980 Gini
Alabama	\$574	21.2	\$18,242	22.9
California	782	14.3	19,076	21.8
Colorado	836	12.1	19,640	19.8
Connecticut	642	13.1	20,145	18.3
Georgia	513	21.3	18,707	22.1
Illinois	633	16.8	20,401	18.6
Indiana	582	15.9	19,633	19.6
Kentucky	573	18.5	19,279	20.2
Louisiana	551	17.7	18,687	22.1
Michigan	569	15.3	20,555	18.3
New Jersey	662	16.1	20,024	19.2
New York	669	15.5	19,041	21.3
Ohio	729	13.0	20,033	18.5
Pennsylvania	618	16.1	19,664	19.3
South Carolina	356	32.5	18,479	22.1
Tennessee	531	20.8	18,376	22.6
Texas	692	16.7	18,312	23.4
Virginia	651	17.9	19,671	20.2
Washington	782	12.2	20,188	18.7
West Virginia	647	17.4	19,478	20.4
Wisconsin	594	14.4	20,434	17.6
ME-NH-VT*	573	13.3	18,610	20.7
Mass-RI	673	12.4	19,588	19.3
Minn-IA-MO-				
Kan-SD-ND**	652	15.7	20,134	18.6
MD-Del	572	17.2	19,945	19.6
Simple Average	626	16.7	19,454	20.2

*Because of grouping of states in the 1900 survey and 1980 Public Use Sample, we have included several multistate groups. In 1900, this group consists of only Maine and New Hampshire. For 1980, data includes Vermont as well.

**South Dakota and North Dakota are only included in 1980 data.

Sources: Based on data from the Eighteenth Annual Report of the Commissioner of Labor (1903) *Cost of Living and Retail Prices of Food* and U.S. Census of Population and Housing: Public-Use Microdata Samples (1983). Note: All data refer to truncated distributions. See text for details.

Both differences are highly significant as measured by a simple t-test. Of course, the persistence of this regional pattern may in itself reflect the regional persistence of underlying factors that influence inequality.

In light of the historic income differences between North and South in the United States, the most obvious starting point for such an investigation is the Kuznets hypothesis. Modern research on the variation in Gini coefficients across states (or other geographic areas) has often emphasized the role of higher average incomes in reducing inequality. Beginning with Kuznets (1955), there is a considerable literature on the effects of income level on inequality (see especially Amos 1988). The general consensus of economists is that income growth from very low levels of income will raise inequality, but then as development continues higher income levels should bring on greater equality as economic opportunities become more widespread. This is in agreement with Kuznets's original hypothesis of an inverted U curve relating inequality and level of income.¹¹ We would suspect that even in 1900 most of the major cities in the country would be developed enough that they would fall on the downward portion of the curve, i.e., the richer ones would be more equal.

This relationship is quite clear in our 1900 sample. The simple correlation in that year between the Gini coefficients (calculated for the "normal" families in principle cities aggregated by state) and mean income was -0.80. In 1980, the correlation was even more pronounced with a value of -0.96.¹²

Before proceeding, we should note that these correlations are undoubtedly influenced by the truncated nature of our income distribution data. In particular, states with a high mean income have a greater proportion of their families excluded from the sample than those with lower means. In poorer states, families are spread throughout the income brackets included in the sample. In the richer states, many of the included families are concentrated in the higher income brackets, and relatively few families are in the lower income brackets.

We can gauge the importance of this truncation effect from performing a sensitivity analysis on the data for each year. In 1980, we recalculated means and Gini coefficients for each state for two alternative income cutoffs, \$40,000 and \$60,000. The base Gini coefficients (using a cutoff of \$30,000) correlated highly with the two new Ginis. The correlation coefficients (across states) between each possible pair was greater than 0.98. Thus, the basic geographic pattern of inequality is not strongly influenced by the truncation phenomenon. However, the correlation between Gini coefficients and mean incomes is somewhat less stable. As mentioned above, that correlation in 1980 for our base case was -0.96. With the \$40,000 cutoff, this "falls" to -0.90, and with the \$60,000 truncation it "falls" to -0.66. For the 1900 data, we performed a similar analysis with a single alternative truncation at \$900 as opposed to \$1,200 in the base case. Again, the Gini

coefficients correlated at better than 0.98, while the correlation between the ginis and the respective mean incomes "rose" from -0.80 in the base case to -0.90 with the lower truncation. In conclusion, while there is little doubt that mean incomes were negatively correlated with inequality in both 1900 and 1980, because of truncation the reported correlation coefficients probably overstate the relationship.

The importance of mean incomes in explaining interstate variations in metropolitan inequality suggests that perhaps the southern states were more unequal in 1900 only because they were poorer. It seems natural to test this hypothesis by doing an analysis of covariance to determine whether southern states were more unequal even after controlling for their lower mean income levels. We do this with a simple regression equation with mean income (in hundreds of dollars) and a southern dummy variable as independent variables. The regional effect still persists.

$$\begin{array}{lclcl} \text{GINI-1900} & = & 31.80 & - 2.61 * \text{INCOME} & + 3.10 * \text{SOUTH} \\ \text{Adj R}^2 = 0.71 & & & (4.75) & (2.86) \end{array} \quad (1)$$

(Figures in parentheses are t-statistics. Note: each \$100 of income is a twelfth of the income scale.)

Thus, the inequality among "normal" metropolitan families of a southern state with about the same level of mean income as those in a northern one was about three Gini points higher. Our corresponding equation for 1980 shows a stronger effect for the income variable and a weaker one for the southern dummy.

$$\begin{array}{lclcl} \text{GINI-1980} & = & 58.39 & - 5.91 * \text{INCOME} & + 0.49 * \text{SOUTH} \\ \text{Adj R}^2 = 0.92 & & & (12.16) & (2.07) \end{array} \quad (2)$$

For 1980, the income variable is measured in units of \$3,100 each. This somewhat odd unit reflects the approximately 31-fold increase between 1900 and 1980 in the nominal average state income. If all incomes in a state rise by the same percentage and there was no structural change, this adjustment factor would leave the income coefficient in the above equation approximately constant. Thus, it would appear that despite the considerable convergence in both mean incomes and the Gini coefficients over time, the relation between the two in 1980 is much steeper than in 1900. This result is quite consistent with the idea that the entire urban system is moving further down Kuznets's inverted U curve, i.e., that the effect of income on relative inequality becomes more pronounced at higher levels of development.

V. A MULTIVARIATE APPROACH

The persistence of a "southern" effect, even after controlling for mean income, suggests that additional variables related to regional character might also be correlated with income inequality. What factors have been emphasized in the cross-sectional analysis of metropolitan inequality? We can take as representative of this body of literature a recent study by Nord (1990), which is perhaps the closest in spirit to our work here since he considers Gini coefficients for husband-wife families in 100 metropolitan areas.¹³ Reviewing the literature on metropolitan inequality, Nord relates the Gini coefficients in these 100 areas to their level of average family income, racial composition, education level, population size, and industrial structure as well as a regional dummy (Nord 1990, 693).¹⁴ Quite similar variables were included in Garofalo and Fogarty (1979) and other earlier studies. We have tried to include each of these in our analysis.

Race

A variable for which the southern dummy might be proxying in both 1900 and 1980 is racial composition. The proportion of non-whites in a state has been found to raise inequality in virtually all modern studies of the subject. Unfortunately, the *Report* gives no information on the non-white proportion of "normal" families, although it does explicitly mention that non-whites were included in the sample. Thus, we are thrown back on using the proportion of non-white families in the population of major cities in each state in 1900. As mentioned above, we do not know the exact group of cities used in the *Report*. Here we use the Census Bureau's list of principal cities with a population of more than 25,000 in 1900. For each state, we calculate a weighted average of the non-white percentages in each of its cities on this list. The simple correlation of the Gini coefficients and proportion non-white is a strong 0.84. To maintain parallelism, we took from the 1980 Census the proportion non-white in the urbanized areas of each state. Thus, even though we could have obtained data on race directly from the public use sample, we attempted instead to match the procedure being applied to the 1900 data.

Education

Studies of urban and statewide income inequality have generally shown that, controlling for mean income, an increase in the average educational level of the population raises inequality. Unfortunately, we do not have 1900 statistics on education comparable to those for 1980. The best we could manage was to com-

pute an illiteracy rate that represents for each state a weighted average of data for principal cities in that state. As we would expect, this measure has a strong positive simple correlation with inequality, 0.75. Illiteracy was considerably higher in the South than in the North, in rich cities than in poor ones, and in areas with large non-white populations as compared to mostly white ones.

We do not have illiteracy data for 1980. In any event, illiteracy in that year had a very different meaning than in 1900. For the more recent period, we conjecture that dropping out of high school is in many ways functionally similar to illiteracy at the beginning of the century. Therefore, we constructed a variable for 1980 (again from the total urbanized area data, rather than from the public use sample itself) that shows the proportion of those older than 25 who have not finished high school. Interestingly, this variable also has a positive relation to inequality when we look at a simple correlation (0.37).

Urban Hierarchy

Recent studies have found that, when controlling for income level, larger metropolitan areas are more unequal than smaller ones (Long, Rasmussen, and Haworth 1977; Nord 1990). For 1900, we determined the proportion of the population in a state's principal cities that lived in cities greater than a million people. This variable in that year was not highly related to inequality in a simple correlation. It should be noted that only three states, New York, Illinois, and Pennsylvania, had cities greater than a million in 1900. For 1980, the million share again showed only weak negative simple correlations.

Industry Structure

Finally, we include a measure of industrial structure in each year. Several studies have found that cities with a high proportion of manufacturing employment tend to have less inequality than those more oriented to trade, finance, and services. The variable included for 1900 is the manufacturing share in total employment of principal cities of greater than 25,000 population. This had a negative simple correlation with inequality of $-.48$. For 1980, the variable is the share of manufacturing employment in urbanized areas.

The basic results are presented in Table 2. For both 1900 and 1980, we find little effect of the manufacturing share on our measure of inequality. It is significant in neither year and has the unexpected positive sign. Thus, we have estimated the equations both with and without this variable. The rest of this discussion focusses on the equations estimated without this variable. The results for the hierarchy variable of a million or more gave more consistent results than

TABLE 2
Determinants of Inequality—Multivariate Equations

Variable	Equation 3 (1900)	Equation 4 (1900)	Equation 5 (1980)	Equation 6 (1980)
Intercept	25.41	29.15	60.054	59.37
Mean	-1.77 (-3.47)	-2.04 (-5.17)	-6.385 (-15.0)	-6.224 (-16.4)
% Non-white	0.36 (4.23)	0.31 (5.34)	0.033 (2.82)	0.031 (2.74)
% Illiterate	-0.65 (-3.46)	-0.58 (-3.46)		
% High school dropouts			-0.046 (-1.70)	-0.030 (-1.71)
% Areas >1Mil	0.03 (2.36)	0.03 (2.25)	0.006 (1.70)	0.005 (1.55)
South	0.32 (0.31)	0.37 (0.37)	0.601 (1.86)	0.407 (2.05)
% Manufacturing	0.05 (0.83)		0.023 (0.77)	
Adj. R2	0.88	0.88	0.96	0.96

Note: The dependent variable for 1900 is the Gini coefficient for "normal" families in the principle cities of each state or state group estimated from the *Report*, while for 1980 the dependent variable is the Gini coefficient for our matched subsample from the Public Use Sample for metropolitan areas grouped by state or state group. See text for discussion.

that for a half million. Thus, what follows focusses only on the equations that include the former variable (Equations 4 and 6).

Looking first at the equation for 1900, we see that even after controlling for the new variables, the mean income level maintains a strong and significant negative effect on inequality. The proportion non-white has a very strong positive effect on inequality. Thus, a one percentage point increase in the proportion non-white in a state raised the 1900 Gini for "normal" families in its principal cities by .31 Gini points.

The illiteracy rate also has a strong effect in 1900 with a negative sign. An increase in the illiteracy rate of one percentage point (the average across state urban areas was 7.8 percent) is associated with a reduction of the Gini coefficient of 0.58 points. It should be kept in mind that these results hold after controlling for the other factors and in particular for mean income.

As hypothesized, the inequality in the large urban sector of the states with major cities populated by more than a million people was statistically greater in 1900. Although significant, the magnitude of the coefficient is of modest size.

As indicated in Equation (4), the regional dummy for the South is virtually insignificant after controlling for this set of variables. This is probably because of the close correlation between region and proportion non-white in 1900.

Since variable definitions and data collection techniques have changed considerably over the years, we cannot rigorously test the hypothesis that the 1900 equation is the same as the 1980 one. Nevertheless, the results presented in Table 2 suggest important differences and similarities.

As noted above, the slope of the mean income variable in 1980 is considerably higher than for 1900. The non-white coefficient is still positive and significant, but its absolute magnitude is only about a tenth of the 1900 value. Similarly, the high school dropout variable, while significant and working in the same direction as the 1900 illiteracy variable, has a much smaller coefficient than the earlier variable. And finally in 1980, the regional dummy for the South looms as relatively more important than the same variable in 1900.

This last observation is perhaps our most interesting. In the earlier year, it seems that the high inequality in Southern metropolitan areas can be adequately explained by pointing toward the region's low-income level and relatively high proportion of non-whites. However, in 1980 when Southern metropolitan areas had mean incomes much closer to those of the rest of the country and the proportion non-white in the Southern areas was also less exceptional, the relatively high inequality in the metropolitan South is not so easily explained. Despite the dramatic fall in the coefficient of variation of the dependent variable, the effect of the Southern dummy actually increases slightly and its significance level goes up dramatically.

VI. CONCLUSIONS

The above statistical results are a first effort to use cross-sectional data to explore the long-term evolution of inequality in the United States. The data set for 1900 is much more restrictive than we would desire. It is difficult (perhaps impossible) to determine whether within the stated guidelines surveyors actually attempted a random sample. The rise in average state inequality for "normal" families between 1900 and 1980 must be approached cautiously. But there is little reason to doubt the reality of the cross-sectional patterns observed.

Essentially the same simple model does a respectable job of explaining variations in inequality in two years separated at such a considerable distance. Inequality

levels across the country have converged a great deal over time as have mean incomes, but the variation that remained in 1980 followed patterns similar to the larger variation of 1900. The persistence of mean income as an independent variable suggests a structural relation that echoes the long-standing claim that prosperity encourages equality.

The continuing importance of race in the determination of inequality is upsetting, but its declining significance is encouraging. This coefficient fell from .31 to .03 (after controlling for the other major variables) in 80 years.

Greater education is generally taken as a spur to greater equality. While our results for both 1900 and 1980 do not definitively argue against such a position, they do suggest that as the larger share of the population receives greater education, an isolated remainder may well suffer relatively, thus increasing inequality. This issue is worth further consideration both in the historical context of 1900 and in the policy context of the present day.

It appears that the importance of industrial structure observed in post-World War II data sets was not a major feature of big-city inequality at the beginning of the century. If this phenomenon is real, its relation to market and institutional changes should prove quite interesting. This point deserves a more ample treatment.

Finally, we find no decline in the significance of region. If anything, we see an increase in the regional effect, at least after controlling for other variables included in the analysis. Inequality in the metropolitan South in 1900 reflected low incomes and racial discrimination. By 1980, the explanatory power of these variables, while still substantial, falls considerably short of the 1900 benchmark. Region in the later year is a more elusive quality. The changing influence of region on metropolitan income inequality invites further study.

ENDNOTES

1. The *Report* itself uses quotation marks when discussing "normal" families.

2. Actually, the question of establishing an appropriate sample plays a secondary role in Williamson and Lindert's criticism of Kravis's discussion of early twentieth century inequality. They are primarily concerned with Kravis's use of estimates for 1910 generated (somewhat mysteriously) by W.I. King. We find Kravis quite cautious in his references to King's estimates. See Kravis (1962, 208).

3. Ideally we would like to have a third year sometime in the middle of our period. Unfortunately, the public use sample for 1940 provides no data on in-

come. For a discussion of changes in state income distributions between 1950 and 1960, see Al-Samarrie and Miller (1967). Sale (1974) covers the period 1950-1970.

4. The precise list of principle industrial centers used in the *Report* is not available. The term "principle cities" was used in the 1900 Census to refer to cities with populations greater than 25,000. However, a similar designation was used to encompass several other groups of cities. An urbanized area in 1980 generally had to have a core city of 50,000 residents.

5. The 1900 figure comes from dividing total family income of \$14.7 billion, as estimated by Kravis (1962, 22), by the total number of families in the 1900 Census, 16.2 million. The latter number includes single-person households, so if anything \$900 would seem to be a lower boundary. The 1980 figure comes from U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 132, *Money Income of Households, Families and Persons in the United States: 1980*, U.S. Government Printing Office, Washington, D.C., 1982.

6. Thus, the sample includes the principal cities or urbanized areas aggregated for each of the following 21 states: Alabama, California, Colorado, Connecticut, Georgia, Illinois, Indiana, Kentucky, Louisiana, Michigan, New Jersey, New York, Ohio, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington, West Virginia, and Wisconsin. In addition, we have observations for the following four groupings: (1) Maine and New Hampshire (1900), Maine, New Hampshire, and Vermont (1980); (2) Massachusetts and Rhode Island (both years); (3) Iowa, Minnesota, Missouri, and Kansas (1900), Iowa, Minnesota, Missouri, Nebraska, Kansas, and North Dakota (1980); and (4) Delaware and Maryland (both years).

7. Using essentially the same data, Kravis reported a Gini coefficient of 16.

8. Kravis uses a somewhat different definition of the sample than we do as dictated by his different data set.

9. Data in the *Report* are given for the industries of 24,000 family heads included in a broader survey. For this larger sample, manufacturing accounts for close to two-thirds of all family heads.

10. On the importance of a geographical reference group in the perception of inequality, see Persky and Tam (1990).

11. But see Amos (1988) for a discussion of a tendency for inequality to again increase at very high levels of development.

12. When we perform a simple regression of Gini coefficients on mean incomes and the square of mean incomes, we find in 1980 that the two independent variables are hopelessly multicollinear. In 1900, we actually find the square term increasing inequality, i.e., a U-shaped curve, not an inverted U curve. However, the magnitude of this effect is quite small, and hence we ignore it in what follows.

Notice that this latter result supports our contention that even for the very poorest cities in 1900 increases in income were likely to reduce inequality.

13. Moreover, his data for 1980 are drawn from the same basic source we use, the 1980 1 in 100 Public Use Sample. But note that to maintain comparability to the 1900 data we have grouped urbanized areas by state and made the further restrictions to match "normal" families described above.

14. Nord also includes a measure of suburbanization, the share of population 18 to 65, and the unemployment rate. The first of these has little relevance to the early part of the century, the significance of the second is greatly limited by our concentration on "normal families" as is the last since we require a male adult in the husband-wife family to be employed.

REFERENCES

- Al-Samarrie, Ahmad, and Herman Miller. "State Differentials in Income Concentration." *American Economic Review* 57 (1967): 59-72.
- Amos, Orley. "Unbalanced Regional Growth and Regional Income Inequality in the Latter Stages of Development." *Regional Science and Urban Economics* 18 (1988): 549-566.
- Danziger, Sheldon. "Determinants of the Level and Distribution of Family Income in Metropolitan Areas, 1969." *Land Economics* 52 (1976): 467-478.
- Eighteenth Annual Report of the Commissioner of Labor. *Cost of Living and Retail Prices of Food*. Washington, D.C.: U.S. Government Printing Office, 1903.
- Garofalo, Gaspar, and Michael Fogarty. "Urban Income Distribution and the Urban Hierarchy-Equality Hypothesis." *Review of Economics and Statistics* 61 (1979): 381-388.
- Kravis, Irving. *The Structure of Income: Some Quantitative Essays*. Washington: University of Pennsylvania, 1962.
- Kuznets, Simon. "Economic Growth and Income Inequality." *American Economic Review* 45 (1955): 1-28.
- Lebergott, Stanley. *The Americans: An Economic Record*. New York: Norton, 1984.
- _____. "The Income Distribution in 1900." In *The American Economy*. Princeton, N.J.: Princeton University Press, 1976.
- Nord, Stephen. "Income Inequality and City Size." *Review of Economics and Statistics* 62 (1980): 502-508.

- _____. "Service Industries, Two-Working Spouse Families and the Distribution of Earnings in Metropolitan Areas." *Review of Economics and Statistics* 72 (1990): 692-696.
- Persky, Joseph, and Mo-Yin Tam. "Local Status and National Social Welfare." *Journal of Regional Science* 30 (1990): 229-238.
- Sale, Tom. "Interstate Analysis of the Size Distribution of Family Income, 1950-1970." *Southern Economic Journal* 40 (1974): 434-441.
- U.S. Census of Population and Housing: Public-Use Microdata Samples Technical Documentation. (Prepared by the Data User Services Division, Bureau of the Census.) Washington, D.C.: The Bureau, 1983.
- Williamson, Jeffrey, and Peter Lindert. *American Inequality*. New York: Academic Press, 1980.