University of Wisconsin-Madison

IRP Discussion Papers

Saul Schwartz

EARNINGS CAPACITY AND THE TREND IN INEQUALITY

DP #757-84

Institute for Research on Poverty
Discussion Paper no. 757-84

Earnings Capacity and the Trend in Inequality

Saul Schwartz
Department of Economics
Tufts University

October 1984

This research was supported in part by a contract to the Institute for Research on Poverty of the University of Wisconsin from the U.S. Department of Health and Human Services. The views expressed in this paper are those of the author and do not necessarily reflect the views of either DHHS or of the Institute for Research on Poverty.

ABSTRACT

This paper updates, revises, and extends the work of Irwin Garfinkel and Robert Haveman in their 1977 monograph, in which they proposed replacement of measured current money income as an indicator of economic status with a measure which they call earnings capacity. Defining earnings capacity as "the potential of the [household] unit to generate an income stream if it were to use its physical and human capital to capacity," they calculate its distribution in 1973.

Despite its well-known drawbacks, traditional programs (such as AFDC) have based aid on current money income. Recent discussions concerning limiting government transfers to the "truly poor" have aroused fresh interest in specifying who are truly poor.

This paper presents two empirical results of interest to the concept of earnings capacity. First, roughly three-quarters of the inequality in earnings is due to differences in earnings capacity. The remainder is attributed to differences in labor supply. This finding is similar to the estimate of Garfinkel and Haveman. Second, the results suggest that inequality in earnings capacity increased from 1967 to 1979, while Garfinkel and Haveman suggest the reverse.

Earnings Capacity and the Trend in Inequality

Saul Schwartz

Recent discussions about concentrating government transfers on the "truly poor" have naturally aroused interest in specifying just who the truly poor might be. Traditional programs (such as AFDC) have based aid on current money income, despite the well-known drawbacks of that measure of economic status.[1]

In their 1977 monograph, Earnings Capacity, Poverty and Inequality, Garfinkel and Haveman (GH) propose replacing measured current money income as an indicator of economic status with a measure which they call earnings capacity. They define earnings capacity as "the potential of the [household] unit to generate an income stream if it were to use its physical and human capital to capacity" (p. 2). Using this definition, GH calculate the distribution of earnings capacity in 1973.

The study by GH, as well as this one, are "in the tradition of efforts to develop a measure of economic status that avoids the inadequacies of the current income indicator" (GH, p. 2). For example, Friedman (1957) noted that consumption (a proxy for economic status) was not as closely related to current income as one might think, and hypothesized that consumption was, instead, a function of "permanent" income. As another example, Weisbrod and Hansen (1977) adjust reported incomes by imputing to each person a stream of income from assets.

If earnings capacity could be successfully measured, it would be a valuable tool for analyzing the welfare implications of any distribution of current money income. Suppose that all the variation in money income was due to labor supply choices. In this case, low money income might not be a justification for government transfers. Suppose, in contrast, that the distribution of earnings capacity was as unequal as the distribution of money income. Then, any argument that the "problems" of poverty and inequality were due to tastes for leisure would be weakened considerably. One of the striking findings of GH is that 80 percent of the inequality in pretransfer income is due to differences in earnings capacity.

This paper has two major purposes. The first is to present a modified version of the GH methodology for calculating earnings capacity. The second is to use the modified methodology to estimate earnings capacity for both 1967 and 1979. This twelve-year time span was a period of rapid increases in federal spending on programs (such as those concerned with job training and education) which were designed, at least implicitly, to increase earnings capacity. Looking at changes in earnings capacity over time gives a sense of the effectiveness of these programs.

Section I describes the work of Garfinkel and Haveman and illustrates the inconsistencies between their theoretical model and the methodology actually used to estimate earnings capacity.[2] Section II outlines alternative methods for measuring earnings capacity which overcome the inconsistencies in the GH methodology. Section III presents empirical estimates of earnings capacity for 1967 and 1979. Conclusions appear in Section IV.

I. SUMMARY AND CRITIQUE OF GARFINKEL AND HAVEMAN

The discussion of the theoretical and empirical models used by GH has two segments. The first emphasizes the long-run equilibrium nature of their earnings capacity concept, describes how GH actually measure earnings capacity, and demonstrates that their empirical methods imply a specific set of theoretical assumptions. In laying out those assumptions, several inconsistencies between their theory and their empirical techniques become apparent. The second segment is a brief discussion of those inconsistencies. The revisions necessary to make the theoretical and empirical specifications consistent are presented in the next section.

It is common in discussions of the determination of observed earnings to assume that those earnings are the outcome of a utility-maximizing labor/leisure trade-off. In the context of measuring earnings capacity, this assumption implies that each person's observed earnings represent a utility-maximizing equilibrium. It is also common to think of observed earnings as having two components--one permanent and the other transitory. Earnings capacity, as formulated by GH, is based only on the permanent component. Earnings capacity is thus a long-run equilibrium concept.

Underlying GH's attempt to measure earnings capacity is the human capital theory of Gary Becker (1967) and the corresponding theoretical estimating models exemplified by Mincer (1974). Specifically, GH assume that the permanent component of earnings is a linear function of a set of variables (X) which measure human capital stocks. These variables include education and experience. A random error term contains both unmeasured human capital stocks and transitory components of earnings. Also included in X are a set of labor-supply dummy variables which indicate how much each person worked in the preceding year. Algebraically, for the ith person,

$$Y = XB + e, \qquad (1)$$

where Y is observed earnings, X is a (1 x k) vector of independent variables, B is a (k x 1) vector of unknown parameters, and e is a zero mean, constant-variance, normally distributed error term which is independent of X.

Having estimated B using ordinary least squares, GH calculate earnings capacity by (1) changing the values of the labor supply variables for each person to "full-time, full-year"; (2) using the new values and the estimated B to compute an estimate of permanent earnings assuming that the individual worked full-time, full-year, estimates which will be denoted hereafter by Y* or by "permanent earnings"; (3) adjusting Y* downward by the factor (50 - W(su))/50, where W(su) is the number of weeks not worked in the relevant year because of sickness, disability or unemployment;[3] and (4) "adding back" to Y* an estimate of the error term from equation (1).

Each of these four empirical steps is based is on a stated or unstated theoretical argument. Step (1) assumes that all differences in permanent earnings except those due to labor supply are also differences in earnings capacity. Labor supply is exogenously determined. For example, two individuals who supply different amounts of labor but are otherwise similar will have the same Y* and thus the same earnings capacity. But a difference in years of education always implies a difference in both earnings capacity and Y*.

By running the regression on all those who worked, GH are assuming that the coefficients apply equally to all individuals, regardless of their observed labor supply. Using the estimated B in step (2) is therefore appropriate in the estimation of Y* for all individuals.

Because earnings capacity is an equilibrium-based concept, any deviations from full-time, full-year work because of unemployment or illness are treated as if they were going to continue over time. This is the apparent rationale for adjusting Y* downward in step (3).

If GH stopped after steps (1) - (3), the variation in Y* would be substantially less than the variation in observed earnings. Even if all human capital investments were measured perfectly, there would still be a variance in earnings due to transitory fluctuations in earnings. Therefore, the variance in permanent earnings would be less than the variance in observed earnings. If human capital investments are not measured perfectly (as assumed by GH), then these unmeasured investments will also appear in the error term of the earnings equation (along with the temporary fluctuations). Therefore, there is even more reason to believe that the variation in Y* will be smaller than the variation in observed earnings. The true measure of earnings capacity would include earnings due to unmeasured human capital differences and exclude earnings due to transitory fluctuations. For this reason, GH "add back" a dollar amount to permanent earnings as calculated in steps (1) - (3). To calculate this dollar amount, GH "draw" a value from a normal distribution which has a mean of zero and a standard deviation equal to the estimated standard deviation of the error term from equation (1).

The inconsistencies between the theory underlying the above procedures and the actual implementation of the procedures are as follows:

1. In theory, the error term in equation (1) contains both transitory fluctuations and unmeasured human capital differences. Because of the unmeasured human capital differences in the error term, Y* is not

equal to earnings capacity. To come closer to true earnings capacity, GH should "add back" to Y* an estimate of only the part of the error term which reflects human capital differences. However, they use an estimate which represents the variation due to both unmeasured human capital and transitory components.

2. Suppose that inconsistency (1) was irrelevant because there were no transitory elements in the error term. Due to unmeasured human capital differences, earnings capacity will still deviate from the Y*, so an estimate of the unmeasured differences should be "added back." But the appropriate standard error to use in "adding back" variation to the Y* is the standard error of full-time, full-year earnings, conditional on X. GH use the standard error of Y in equation (1), where the sample includes all workers, not just full-time, full-year workers. This standard error is much larger than the corresponding standard error for only full-time, full-year workers (see Tables 1 and 2).

3. The coefficients in equation (1) are applied to all individuals (since labor supply is assumed to be exogenously determined). However, if labor supply is really endogenous, then the estimates of B will be biased.[4]

4. In theory, earnings capacity is a long-run equilibrium concept. By adjusting Y* (and thus earnings capacity) downward to reflect the number of weeks not worked due to unemployment (rather than choice), GH are assuming that unemployment and disability are characteristic of the long-run equilibrium. But it is more reasonable to follow the macroeconomic practice of assuming that there is no unemployment in a long-run equilibrium.

Table 1

Earnings Functions for Black Men
1967, 1973 and 1979 CPS Data

Independent Variable	Dependent Variable: Natural Logarithm of Earnings		
	1967	1973	1979
	(1)	(2)	(3)
Years of schooling	.00391	-.0088	.0221
(Years of schooling)2	.00248	.0017*	.00242*
Age	.0670*	.0525*	.0744*
(Age)2	-.000863*	-.0007*	-.000793*
Age x yrs schooling	.0000601	.0004	-.000219
Weeks worked:			
1-13	-3.017*	-2.0173*	-2.067*
14-26	-.595*	-.8324*	-.945*
27-39	-.369*	-.3742*	-.586*
40-47	-.212*	-.2563*	-.192*
48-49	-.0993	-.0970	-.0324
50-52	---	---	---
Full-time work	---	---	---
Part-time work	-.905*	-.9827*	-.972*
Location:			
Northeast	-.129	-.0197	-.107
South	-.353*	-.2362*	-.190*
West	-.120	.0132	-.136
SMSA suburb	.301*	.2664*	.163*
SMSA central city	.246*	.1609*	.162*
Non-urban	---	---	---
Constant	6.915*	7.6699*	7.355*
R-squared	.560	.607	.590
Adjusted R-squared	.557	---	.587
F-statistic	208.18	266.86	219.58
Sample size	2637	---	2462
Mean of ln(earnings)	8.016	---	9.085
Standard error of regression	0.89075	---	0.62588

Source: Columns (1) and (3): Computation by author from CPS data supplied by the Institute for Research on Poverty; Column 2: Garfinkel and Haveman, (1977) Earnings Capacity, Poverty and Inequality, pp. 12-13.

*Significantly different from zero at the 0.01 level of significance.

Table 2

Earning Functions for
Full-Time, Full-Year Workers
CPS Data, 1967 and 1979

Independent Variable	Dependent Variable: Natural Logarithm of Earnings			
	Uncorrected		Corrected	
	1967	1979	1967	1979
	(1)	(2)	(3)	(4)
Years of schooling	.0129	.0108	.0141	.0168
Years of schooling2	.00267*	.00284*	.00214*	.00234*
Age	.0382*	.0607*	.0144*	.0384*
Age2	-.000374*	-.000607*	-.000074*	-.000331*
Age x Yrs schooling	-.000571	-.000334	-.000644	-.000475
Weeks worked:				
1-13	---	---	---	---
14-26	---	---	---	---
27-39	---	---	---	---
40-47	---	---	---	---
48-49	---	---	---	---
50-52	---	---	---	---
Full-time work	---	---	---	---
Part-time work	---	---	---	---
Location:				
Northeast	-.0809*	-.0863	-.0810*	-.0876
South	-.261*	-.120*	-.265*	-.122*
West	-.00744	-.0216	-.0126	-.0128
SMSA suburb	.194*	.171*	.193*	.167*
SMSA central city	.144*	.108*	.144*	.113*
Non-urban	---	---		
Selectivity bias cor.	---	---	.132	.0930
Constant	7.454*	7.742*	7.676*	8.044*
R-squared	.293	.241	.299	.246
Adjusted R-squared	.289	.236	.295	.241
F-statistic	70.29	47.73	65.87	44.61
Sample size	1707	1516	1707	1516
Standard error of regression	0.36544	0.41200	0.48984	0.46245

Source: Computations by the author from CPS data provided by the Institute for Research on Poverty.

*Significantly different from zero at the 0.01 level of significance (but see Appendix A for a discussion of the standard errors).

**The reported standard error has been corrected for the bias introduced by correcting for selectivity bias. See Appendix A for a discussion of the procedure used to calculate these standard errors.

II. REVISIONS TO THE EARNING CAPACITY METHODOLOGY

Interpreting earnings capacity as a long-run equilibrium measure of economic status requires several modifications to the GH methodology. First, the crucial variable in the measurement of earnings capacity, on both the theoretical and empirical levels, is the error term from equation (1). GH provide conflicting accounts of its theoretical composition. Sometimes it contains only transitory fluctuations in earnings; other times it contains only unmeasured human capital differences. In still other instances, it is a melange of those two variables and tastes for leisure.

Interpreting earnings capacity as a long-run measure resolves the confusion about the nature of the error term in equation (1). In a long-run equilibrium, there are no transitory fluctuations, so I will assume that it consists only of unmeasured variation in human capital stocks. Given this assumption, the standard error of full-time, full-year earnings becomes the appropriate basis for "adding back" a dollar amount to Y* calculated from equation (1).

The appropriate variation to be "added back" is the variation in full-time, full-year earnings. It is not the variation in the error term in equation (1), since that variation applies to all workers. Therefore equation (1) is estimated on a sample of full-time, full-year workers. This sample selection will not only produce the appropriate standard error but will also avoid the bias created by the inclusion of the labor supply variables on the right-hand side of the equation. Of course, selecting a sample on the basis of an endogenous variable creates a selectivity bias, which if uncorrected would affect the estimates of B. I correct for selectivity bias using methods described in Appendix A.

Second, a typical macroeconomic assumption is that there is no involuntary unemployment in a long-run equilibrium. Therefore, it is not necessary to adjust earnings capacity downward by a factor reflecting the amount of time not worked because of involuntary unemployment.

I make two other modifications which are theoretically consistent, but have smaller impacts on the results. These involve using the CPS weights in estimating earnings capacity and "adding back" variance in a way different from the GH method. These modifications will be discussed where appropriate.

III. EMPIRICAL METHODS AND RESULTS

In this section, earnings capacity is estimated using the modified methodology, and the results are compared to earnings capacity estimated using the original methods of GH.[5] They estimated earnings capacity in 1973, but it is important to measure earnings capacity and its distribution in 1967 and 1979, since many programs designed to increase earnings capacity for various groups were implemented in that period. These changes in the distribution of earnings capacity (and the methodological changes) are illustrated here by examining the changes in the observed earnings and earnings capacity of black men.

For the purposes of comparing the modified methodology to the original GH methodology, Table 1 shows GH-style earnings functions of black men for 1967 and 1979. The coefficients are broadly similar. The familiar parabolic relationship between earnings and age can be seen in all three years. The age-earnings profile is flattest in 1973, but roughly comparable in 1967 and 1979. Schooling is positively, but not linearly, related to earnings. Of the three "years of schooling" variables, only the coefficients on squared years of schooling are significantly different from zero, suggesting that the effect of additional years of schooling on earnings is exponential. The coefficients are similar in magnitude across the years.

As far as the "region of residence" variables are concerned, only the coefficient on South is significant in all three years. It declines in absolute magnitude over time, perhaps indicating rising relative wages in the "New South." Living in an SMSA (whether in a suburb or in the central city) becomes less of an advantage over time, although the advantage is significant in all three periods.

As we would expect, the coefficients on the labor supply variables (where full-time, full-year is the excluded category) are negative, very large in absolute value, and significantly different from zero. Also as expected, the coefficients decline in absolute magnitude as labor supply increases.

The regression statistics are similar in all three years. Given the cross-sectional nature of the regression, a very high percentage of the total variation (around 60 percent) is explained. However, the estimated standard errors of the regression are also high (0.6 to 0.9), a fact which will play a role in the "adding back" of variance to earnings capacity estimates.

The first modification is to include only full-time, full-year workers in the earnings capacity equation. However, because there may be unobserved characteristics by which individuals "select" themselves as full-time, full-year workers, assigning earnings capacity to part-year or part-time workers based on a regression using a sample of full-time, full-year workers will be inaccurate. While this problem can not be dealt with in an entirely satisfactory way, it has become relatively common to include an additional variable in earnings (or wage) regressions to attempt to correct for this selectivity bias. The relevant model, adapted from Heckman (1979), is described in Appendix A.

Table 2 shows the earnings functions estimated only for those who worked full-time, full-year. The coefficients in columns (1) and (2) are uncorrected for selectivity bias while those in columns (3) and (4) are corrected. The correction makes a difference in both years, as indicated by the larger coefficients on education and by the flatter age-earnings profiles. However, the other coefficients remain roughly constant. The coefficients on the selectivity bias correction variables are positive and significantly different from zero. The discussion below refers to columns (3) and (4).

In principle, there is no reason for the explanatory power of the model in Table 1 to be greater than that of Table 2. But in the regressions in Table 1, the R-squared is 0.560 in 1967 and 0.590 in 1979, while in the regressions of Table 2, the R-squared is 0.295 in 1967 and 0.242 in 1979. This cannot be explained simply by the exclusion of the labor supply variables from the regressions in Table 2, since the variation in the dependent variable has also decreased because of the sample restrictions. The fact that the R-squared drops by so much suggests that the labor supply, variables explain a great deal of the (greater) variance in earnings in the sample which includes all workers.

There are both similarities and differences in the coefficient estimates in the regressions of Tables 1 and 2. The coefficient on education (entered linearly) remains insigificant. The coefficients on squared years of education remain significantly different from zero and have approximately the same magnitude. The biggest difference is in the estimated age-earnings profile. The coefficients in Table 1, columns (1) and (3), indicate that the age-earnings profile was approximately the same in both 1967 and 1979. However, the coefficients in Table 2 suggest that the age-earnings profile for black men became much more peaked over the time period. The location variables have the same pattern of signs in both sets of regressions, although the magnitude of the coefficients is smaller in the second set.

Even though the explanatory power of the model in Table 2 is lower as measured by the R-squared, the estimated standard errors are considerably lower. Comparing the regressions using 1967 data, the standard error in Table 2 is about 55 percent of that in Table 1. Using 1979 data, the standard error from Table 2 is about 75 percent of the standard error in Table 1. As will be discussed later, this standard error is the critical variable in estimating earnings capacity. The lower the standard error, the less variance is "added back" to estimates of permanent earnings.

Given the set of coefficients from an earnings function, the next step in estimating earnings capacity and its distribution is to assign a full-time, full-year permanent earnings (Y*) to each individual. The modified regression specification demands a different procedure for calculating earnings capacity than that used by GH and outlined at the beginning of Section I.

Among those who are included in the sample for the earnings regressions, permanent earnings is the fitted value of the regression (columns (3) and (4) from Table 2). For all those excluded from that regression--anyone who did not work full-time for 50-52 weeks in the previous year--permanent earnings is calculated by using the earnings function coefficients and the relevant individual characteristics. The variable representing the correction for selectivity bias is not used in the imputation, for reasons discussed in Appendix A. No downward adjustment is made for unemployment, and permanent earnings is set equal to zero for anyone who did not work at all in the relevant year because of illness or disability.

In addition, there are two other modifications which I make to the GH methodology. The first of these is to utilize the population weights in order to avoid any implication that earnings capacity can be attributed to individuals. The use of the weights also allows me to make a second modification--a different procedure for "adding back" variance, described below.

The first column of the top two panels of Table 3 shows the distribution of Y* using the GH style regressions from Table 1 and their adjustment method, described in Section I. The lower two panels show the distribution of permanent earnings using the regressions of Table 2 and the modified method of adjusting individuals up to full-time, full-year work. The second column of Table 3 shows the distribution of observed earnings, weighted and unweighted.

First, note that the sample sizes used in constructing Table 3 (line 2 in each Panel) are different from those in Tables 1 or 2. This reflects the fact that Y* is estimated for everyone in the data set, including those who were excluded from the earnings regressions of Tables 1 and 2.

Table 3

Distributions of Earnings Capacity and Observed Earnings, 1967 and 1979: No Variance "Added Back" to Fitted Values

	Fitted Earnings	Observed Earnings
	Garfinkel-Haveman Earnings Functions (Unweighted 1967 Data)	
Mean	$3,791	$3,907
Number of cases	3,037	3,037
Total dollars (mill.)	$11,513	$11,866
Gini coefficient	0.31121	0.42397
	Garfinkel-Haveman Earnings Functions (Unweighted 1979 Data)	
Mean	$10,408	$9,443
Number of Cases	3,232	3,232
Total Dollars (mil.)	$33,639	$30,520
Gini Coefficient	0.29047	0.49602
	Modified Earnings Functions (Weighted 1967 Data)	
Mean	$3,989	$3,975
Number of Cases (000)	3,983	3,983
Total Dollars (mil.)	$15,889	$15,836
Gini Coefficient	0.16402	0.41838
	Modified Earnings Functions (Weighted 1979 Data)	
Mean	$10,048	$9,660
Number of cases (000)	4,981	4,980
Total dollars (mil.)	$50,051	$48,118
Gini coefficient	0.20466	0.48497

Source: Computations by the author from data provided by the Institute for Research on Poverty.

Second, note that with one exception, the mean of permanent earnings is higher than the mean of the observed values. This is because the permanent earnings of those who do not work full-time, full-year is higher than their observed earnings.[6] Third, in the upper two panels, the relationship between the Gini coefficients[7] for the distribution of permanent earnings in the two years shows the opposite pattern from those for actual earnings. The distribution of observed earnings became more nequal, while the distribution of permanent earnings became more equal. Using the modified procedures, the distribution of both observed earnings and permanent earnings become more unequal.

Last, note that in all cases the Gini coefficient for permanent earnings is substantially lower than that for observed earnings. The ratio of the two coefficients ranges, with one exception, from about .39 to about .59. If permanent earnings were the same as earnings capacity, then we would have to attribute only 40 to 60 percent of the inequality in the distribution of income to variation in earnings capacity and the remainder to labor supply choices.[8]

In order to assume that permanent earnings are the correct measure of earnings capacity, the error terms of the regressions in Tables 1 and 2 must consist only of transitory fluctuations in earnings. However, it is likely that the error terms also contain unmeasured human capital differences.

GH recognize that part of the error term is attributable to unmeasured human capital differences. They write "To the extent that [the error term] is attributable to unobserved human capital differences or to chance, its suppression is inappropriate for many purposes. To avoid this artificial compression of the earnings capacity distribution, we distribute individual observations ... about the ... mean"(GH, p. 15).

However, to the extent that the error term consists of chance elements, this "adding back" of variance is itself incorrect. Ideally, we would like to be able to decompose the error term into a part due to differences in capacity and a part due to transitory or chance factors. Lacking such a decomposition of the error term, it is in keeping with the long-run spirit of the analysis to assume that e(i) is composed entirely of unmeasured differences in earnings capacity. If so, then an estimate of e(i) should be "added back" to the fitted value in order to obtain a better estimate of earnings capacity. This is, in fact, what GH do without making a consistent set of assumptions about e(i).

In order to form an estimate of the error term for each person, GH use the assumption of the classical linear regression model that, in the population, observations on the dependent variable are distributed normally around the regression line for any given set of independent variables. Therefore, they "draw" a value of e(i) for each person from a normal distribution with a mean of zero and a standard deviation equal to the standard error of the regression reported in Table 1. Using a random number generator, they assign each individual a single estimated e(i) and add it to Y* to compute earnings capacity.

The correct way to "add back" variance (assuming thatthat variance to be added back is indeed the standard error of e(i)) is to utilize the population weights (W) given in the CPS data. These weights represent the number of individuals in the general population who are observably identical to the sample individuals in terms of age, race and sex. Since the assumption is that earnings in the population are distributed normally around the regression line, conditional on X, a separate normal distribution can be created for each person in the sample. This normal distribution indicates the distribution of earnings of the W people in the population corresponding to the fitted value for each sample person. The variance of this normal distribution is, by the assumption of homoscedasticity, the same for all people. An estimate of that common variance is the square of the standard error of the regression. For example, suppose that the relevant standard error is 0.9 and consider a sample person with a fitted value of 9.0. Suppose further that the relevant population weight is 1000. If earnings in the population are normally distributed about 9.0, then we know that 3.83 percent (38.3 of the 1000) individuals have a logarithm of earnings between 9 and (9 + 0.1(0.9)), or 9.09. This is because 0.0383 of the area under a normal distribution lies between the mean and a point which is 0.1 standard errors above the mean.

Column (1) of Table 4 is simply the actual distribution of earnings for black men, calculated using the population weights contained in the CPS data. If an individual's CPS weight is W, then the calculations underlying this column assume that there are W individuals in the population with exactly the same earnings as the sample individual. They are the same as the Gini coefficients reported in Column (2) of Table 3. As noted there, the distribution of observed earnings became more unequal (Column (1)).

Column (2) of Table 4 represents my estimate of the distribution of earnings capacity in the population. Comparing columns (1) and (2) for each year shows the decomposition of the distribution of earnings into a

Table 4

Distributions of Earnings Capacity and
Observed Earnings, 1967 and 1979
Variance "Added Back" to Permanent Earnings

	Observed Earnings (Weighted) (1)	Earnings Capacity (Weighted) (2)
	Earnings Functions from Table 2 (Weighted 1967 Data)	
Mean	$3,975	$3,989
Number of cases (000)	3,983	3,983
Total dollars (mill.)	$15,836	$15,889
Gini coefficient	0.41838	0.33429
	Earnings Functions from Table 2 (Weighted 1979 Data)	
Mean	$9,660	$11,250
Number of cases (000)	4,981	4,981
Total dollars (mill.)	$48,118	$56,037
Gini coefficient	0.48597	0.35851

part due to earnings capacity and a residual part which is assumed to be due to labor supply choices or "capacity utilization."

The results in Table 4 suggest that for black men, the proportion of the inequality in the distribution of earnings which is due to differences in earnings capacity is about 80 percent in 1967 (0.33/0.42) and 75 percent in 1979 (0.36/0.49). This estimate is comparable to the GH estimate of 80 percent for 1973.

Another important result from Table 4 is that the distribution of earnings capacity became more unequal, as did the distribution of observed earnings. However, the increase in inequality is relatively small for earnings capacity as compared to observed earnings. This suggests that the increase in inequality over the time period was primarily due to changes in labor supply choices rather than changes in earnings capacity.

Mean income in column (2) is higher than mean income in columns (1). This is because all those individuals who had low earnings in the observed distributions (because they did not work full-time, full-year) have been assigned the full-time, full-year earnings of those with exactly their observed characteristics. That is, the variable whose distribution is being considered is earnings capacity, not observed earnings, and mean earnings capacity should be higher than mean earnings.

To gauge the impact of the methodological modifications implemented here, it is useful to compare the results of Table 4 to similar results using the original GH methodology. The modified methodology yields Gini coefficients for earnings capacity in 1967 and 1979 of .33 and .36 (Table 4). The comparable Gini coefficients using the GH methodology are .56 and .48 (Appendix Table B.1). So not only does the modified methodology yield dramatically smaller Gini coefficients, but the direction of change is also different. Furthermore, the GH methodology implies a distribution of earnings capacity in 1967 which is actually more unequal than the distribution of observed earnings. A more complete description of the results using the GH methodology appears in Appendix B.

V. SUMMARY AND CONCLUSIONS

The purpose of this paper has been to reexamine the earnings capacity methodology developed by Garfinkel and Haveman in order to use it to compare the distributions of income in 1967 and in 1979.

This paper has reviewed the work of Garfinkel and Haveman and constructed a theoretical framework consistent with the empirical methods employed. The empirical work began by implementing some changes in the measurement of earnings capacity, making that measurement consistent with the long-run equilibrium focus of earnings capacity. The result was a set of earnings capacity estimates which are better not only in theory but in the sense that a key estimate--the standard error of the regression--is better. In estimating the distribution of earnings capacity, I use a method of "adding back" variance which is again theoretically superior and which yields reasonable results.

There are two empirical results of interest. First, I estimate that roughly three-quarters of the inequality in earnings is due to differences in earnings capacity. The remainder is attributed to differences in labor supply. This finding is similar to the GH estimate. Second, the results from my methods suggest that inequality in earnings capacity increased from 1967 to 1979, while the GH methods suggest the reverse.

These results suggest that the programs of the 1970s may have had positive effects on earnings capacity but may also have led to changes in labor supply which made the distribution of earnings more unequal.

APPENDIX A

In calculating earnings capacity, I need to estimate how much an individual would earn if he worked full-time, full-year. My starting point is to assume that, for all individuals, earnings are a function of a 1 x k vector of exogenous variables and a random error term. The exogenous variables include age, education and some demographic variables. Algebraically,

$$Y(1) = XB + e(1), \qquad (A.1)$$

where Y(1) is full-time, full-year earnings, X is the 1 x k vector of independent variables, B is a k x 1 vector of unknown parameters, e(1) is a random error term whose distribution will be discussed shortly, and N is the size of a randomly chosen sample.

If I select a subsample of only full-time, full-year workers, I open the possibility of introducing bias into the estimation of B.[8] The problem can also be thought of in terms of a censored sample, in which observations on X are available for a complete random sample, but observations on Y (full-time, full-year earnings) are available only for a nonrandom subset of observations.

This problem is common to many different areas of empirical research and has been discussed extensively in recent years following the path-breaking work of Heckman (1979). I use a relatively simple version of Heckman's correction for selectivity bias.

I define full-time, full-year workers as those who worked full-time for 48 or more weeks. Therefore, I select a sample of individuals for whom Y(2) > 48 where Y(2) is the number of weeks worked. Suppose that

Y(2) is also determined by X and an error term e(2). That is,

$$Y(2) = XG + e(2) \tag{A.2}$$

where G is a k x 1 vector of unknown parameters.

The joint distribution of the pair or error terms e(1) and e(2) is normal, with zero mean and a covariance matrix

$$S = \begin{bmatrix} s(11) & s(12) \\ s(21) & s(12) \end{bmatrix}$$

The error terms are uncorrelated across observations but not across equations. The variance of e(2) is not estimated and is assumed to equal unity.

Given my restriction of the sample to those who work full-time, full-year, the model of equation (A.1) has the population regression function:

$$E[Y(1) \mid X,Y(2)>48] = XB + E[e(1) \mid X,Y(2)>48] \tag{A.3}$$

In general, the last term in equation (A.3) is nonzero and its omission from the regression will lead to biased estimates of B.

Heckman shows that

$$\begin{aligned} E[e(1) \mid X,Y(2)>48] &= E[e(1) \mid X,XG + e(2) > 48] \\ &= E[e(1) \mid X,e(2) > -XG'] \\ &= s(12)q \end{aligned} \tag{A.4}$$

where q = f(XG')/[F(XG')] and f and F are the density and distribution functions of the standard normal distribution. The vector G' is the parameter vector G with the constant 48 absorbed into the constant term. Therefore, the population regression function for the selected sample can be written:

$$E[Y(1) \mid X,Y(2)>48] = XB + s(12)q \tag{A.5}$$

The inclusion of q as a regressor in the ordinary least squares regression of Y(1) on X will yield consistent estimates of B.

To estimate q, I must first estimate G'. This is done by specifying a dichotomous variable D which is equal to 1 if a person works full-time, full-year, and is equal to 0 otherwise. That is:

$$D = \begin{cases} 0 \text{ if } e(2) < -XG' \\ 1 \text{ if } e(2) > -XG' \end{cases} \qquad (A.6)$$

Consistent estimates of G' can be derived using probit analysis. Estimates of G' appear in Table A.1.

Using the coefficients in Table A.1, the selectivity bias correction factor q is estimated for each individual. It is then included as an independent variable in the regressions reported in columns (3) and (4) of Table 2 in the text.

In imputing earnings capacity, using the coefficients reported in Table 2, there is a question of how to use the correction factor q. If the labor supply status is known, then q must be included in the fitted values calculated from Table 2 since, for example, the population regression function is XB + s(12)q for someone known to be working full-time, full-year.

However, as discussed in the text, I use the weights reported in the CPS in order to avoid the appearance of being able to calculate earnings capacity for any individual. These weights are based only on age, race, and sex so that the labor force status of the group which the individual represents is unknown. Therefore, q is equal to zero for the group in the population since the expected value of e(1) for someone whose labor supply status is unknown is zero. As a result the term s(12)q is not included in calculating the fitted values from Table 2.

When q is included as a regressor in the earnings equations, the conventionally calculated standard error of the regression is biased downward. The standard errors of the coefficients are also biased. For the purposes of this paper, the bias in the standard errors of the coefficients is not important. But the bias in the standard error of the regression is critical to the "adding back" of variance to the estimates of permanent eranings in Table 4. The correct calculation of the standard error was done in the manner suggested by Heckman (1979) and Greene (1980). Denoting the consistent standard error by S^2:

$$S^2 = S^2 - (\;\;) *L$$

where S^2 is the standard error of the regression computed in the conventional way, C is the coefficient on q in the earnings equation and L is the mean of [-q(XG' + q]. The actual values reported in Table 2 are:

For 1968, $(0.48984)^2 = (0.36387)^2 - (0.13240)^2 * (-6.134)$

For 1980, $(0.46245)^2 = (0.41072)^2 - (0.09298)^2 * (-5.225)$

Table A.1

Estimates of the Determinants of
Working Full Time, Full Year

Dependent Variable: 1 If Person Works Full Time, Full Year, 0 otherwise.

Independent Variables	1967 (1)	1979 (2)
Years of schooling	-0.0214	0.0077
(Years of schooling)^2	0.00097	0.00015
Age	0.101	0.164*
(Age)^2	0.00014	-0.0021*
Age x yrs schooling	0.00093	0.0012
Marital status (Married=1)	-0.325*	0.354*
Number of dependents	-0.0091	-0.023
Wealth[a] ($000)	-0.058	-0.058*
Constant	-1.597*	-3.399*
Chi-square[b]	545	896

*Significantly different from zero at the 0.01 level of significance.

[a]Wealth is measured as property income plus interest and dividends.

[b]The null hypothesis for this chi-square test is that all coefficients are zero. There are nine degrees of freedom in each model.

APPENDIX B

Table B.1 shows the distribution of earnings capacity using the GH method of estimating Y* and the GH method of "adding back" variance, described in the text. Notice that in 1967 the distribution of earnings capacity is actually more unequal than the distribution of earnings (0.56 to 0.42). Using the 1979 regression, the Gini coefficient for earnings capacity is almost the same as the Gini for observed earnings (0.48 to 0.50), suggesting that most of the inequality of earnings is due to capacity differences, not labor supply choices.

Comparing the Gini coefficients for 1967 to those for 1979, the distribution of earnings capacity became more equal in contrast to the distribution of observed earnings, which became more unequal.

Last, note that actual 1967 mean earnings were $3,907 while the simulated mean earnings capacity is $5,475. Actual 1979 mean earnings were $9,443, while the simulated mean earnings capacity was $13,085. Because the dependent variable and the estimates of e(i) are in logarithms, the addition of the randomly chosen e(i) to the fitted values does not leave the overall mean unchanged.

Table B.1

Distributions of Earnings Capacity and Observed Earnings, 1967 and 1979; Variance "Added Back" to Fitted Values According to the GH Method

	Earnings Capacity	Observed Earnings
	Garfinkel-Haveman Earnings Functions (Unweighted 1967 Data)	
Mean	$5,475	$3,907
Number of cases	3,037	3,037
Total dollars	$16,629	$11,866
Gini coefficient	0.56113	0.42397
	Garfinkel-Haveman Earnings Functions (Unweighted 1979 Data)	
Mean	$ 1,3085	$9,443
Number of cases	3,232	3,232
Total dollars	$42,289	$30,520
Gini coefficient	0.47603	0.49602

Source: Computations by the author from CPS data provided by the Institute for Research on Poverty.

NOTES

[1]Current money income may be an inadequate measure of long-term economic status for a number of reasons, these including the following: (a) current income is the net result of an optimizing labor-leisure trade-off, and individuals with the same income may work very different amounts of time in order to earn that income; (b) current income may have a large transitory component, reflecting temporary business-cycle conditions or one-time gains and losses; (c) current income must typically be used to support varying numbers of individuals--two families may have the same income but markedly different demographic compositions.

[2]My criticisms of Garfinkel and Haveman should be taken as an effort to build on their work. Measuring what seem to be straightforward theoretical notions of economic status is extremely difficult, and Garfinkel and Haveman made a pioneering effort in that regard.

[3]GH also make an adjustment to reflect costs of working. The most important of these costs is "the need to provide care for children." If all adult members of a household were to work full time, the household would have to pay for child care. The earnings capacity of women in households with minor children is therefore adjusted downward to reflect this cost. The adjustment is ignored for the remainder of this paper, since I deal only with men.

[4]GH acknowledge the potential endogeneity of labor supply in footnote 2 on page 10 of their monograph.

[5]By "original methods," I mean only steps (1) to (4) on p. 4. In their monograph, GH apply their methods not only to black men but also to whites and females and compute household earnings capacity. They also make the adjustment described in footnote 3, above.

[6]The one exception is due to the way Y* is calculated for those who do not work at all owing to unemployment or disability. In the estimates using the GH method, an earnings capacity of zero is assigned to these individuals. If they are excluded from the sample entirely, the mean of Y* is greater than the mean of observed earnings. Furthermore, the Gini coefficients in Table 3, panels 1 and 2, are 0.24370 and 0.17918 respectively.

[7]When I exclude those who did not work at all (see note 6), the Gini coefficients in Table 4, column (1), are 0.52035 and 0.39757, respectively.

[8]I make the sample selection in order to avoid the bias introduced by including the endogenous labor-supply variables as exogenous variables in the earnings equation.

REFERENCES

Becker, Gary. 1967. Human Capital. New York: National Bureau of Economic Research.

Friedman, Milton. 1957. A Theory of the Consumption Function. New York: National Bureau of Economic Research.

Garfinkel, Irwin and Robert Haveman (with the assistance of David Betson). 1977. Earnings Capacity, Poverty, and Inequality. New York: Academic Press.

Greene, William. 1980. Sample selection Bias as a Specification Error: Comment. Econometrica. 47: 153-161.

Heckman, James J. 1979. Sample selection Bias as a Specification Error. Econometrica. 47: 153-161.

Mincer, Jacob. 1974. Schooling, Experience and Earnings. New York: National Bureau of Economic Research.

Weisbrod, Burton and W. Lee Hansen. An Income - Not Worth Approach to Measuring Welfare. In Improving Measures of Economic Well-Being, ed. Marilyn Moon and Eugene Smolensky. New York: Academic Press.