Wage Discrimination: Reduced Form and Structural Estimates

Author(s): Alan S. Blinder

Source: *The Journal of Human Resources*, Autumn, 1973, Vol. 8, No. 4 (Autumn, 1973), pp. 436-455

Published by: University of Wisconsin Press

Stable URL: https://www.jstor.org/stable/144855

WAGE DISCRIMINATION: REDUCED FORM AND STRUCTURAL ESTIMATES*

ALAN S. BLINDER

ABSTRACT

Regressions explaining the wage rates of white males, black males, and white females are used to analyze the white-black wage differential among men and the male-female wage differential among whites. A distinction is drawn between reduced form and structural wage equations, and both are estimated. They are shown to have very different implications for analyzing the white-black and male-female wage differentials. When the two sets of estimates are synthesized, they jointly imply that 70 percent of the overall race differential and 100 percent of the overall sex differential are ultimately attributable to discrimination of various sorts.

It is well known that, in the American economy, whites earn much higher wages than blacks and males earn substantially higher wages than females. In view of the important role which inequality in wages plays in explaining overall inequality of incomes,[1] it seems likely that the white-black and male-female wage differentials are major contributors to the unequal distribution of income in the United States. An understanding of these differentials, therefore, is vitally important where questions of income redistribution are concerned.

The author is Assistant Professor of Economics, Princeton University.

* I owe a great deal to Robert E. Hall who furnished me with the data, assisted with computational problems, and commented astutely on an earlier draft. I am also indebted to Orley Ashenfelter, Yoram Ben-Porath, Nicholas Barr, Ronald Oaxaca, Ray Fair, A. P. Thirlwall, and a referee for their helpful comments. Partial support under National Science Foundation Grant GS 32003X is gratefully acknowledged. [Manuscript received August 1972; accepted December 1972.]

1 Other research I have done suggests that roughly half of total income inequality, as measured by the Gini ratio, is attributable to unequal wage rates. See [2].

It is known that part of each wage differential is due to differences in "objective" characteristics such as education and work experience, while part remains even when white-black and male-female differences in these traits are controlled for. However, the quantitative dimensions of the various causes of unequal wages are not well known. How much of the white-black wage differential is attributable to the superior education of the whites? How much of the male-female differential is due to the fact that men have easier access to the high-paying occupations? It is questions such as these which this paper seeks to answer.

Section I below outlines the techniques I shall use to decompose the two wage differentials into their component causes, a simple technique which is of quite general applicability. Section II discusses the unique features of the data source upon which this study is based. The distinction between reduced form and structural estimation, which is almost always ignored in micro wage equations of the kind employed here, is drawn, and methods for estimating each equation are discussed. Sections III and IV report the results of applying these procedures to the wage differentials between white males and black males, and between white males and white females, respectively.[2] An Appendix provides the regression estimates upon which the results in Sections III and IV are based.

The main findings of this study are two: First, while the overall white-black and male-female differentials are strikingly similar is size, decomposition shows that the qualitative nature of race and sex differentials differ quite radically. Second, whether we use structural or reduced form estimates of the wage equation greatly colors our view of the wage differentials. For example, when we use the structural equations, the regressions suggest that about 60 percent of the white-black wage differential is due to the whites' superior endowments of various characteristics, leaving 40 percent to be attributed to discrimination of various kinds. By contrast, the reduced form estimates attribute only about 30 percent of the raw differential to objective traits and 70 percent to discrimination.

I. THE DECOMPOSITION OF WAGE DIFFERENTIALS

The most common way to study the dispersion in individual wages or incomes is to estimate a regression like:

$$Y_i = \beta_o + \sum_{j=1}^{n} \beta_j X_{ji} + u_i \tag{1}$$

where Y_i is the level or natural logarithm of earnings, income, or wage rate,

2 Due to the small sample sizes, no attempt was made to compare black males with black females or white females with black females.

and $X_{1i}, \ldots, X_{ni}$ are n observable characteristics used to explain Y. If we are particularly interested in comparing two demographic groups (such as whites and blacks), it makes sense to estimate an equation like (1) for each group:

$$Y_i^H = \beta_o^H + \sum\nolimits_{j=1}^{n} \beta_j^H X_{ji}^H + u_i^H \tag{2}$$

$$Y_i^L = \beta_o^L + \sum\nolimits_{j=1}^{n} \beta_j^L X_{ji}^L + u_i^L \tag{3}$$

where the H superscript indicates the high-wage group (always white males in this study) and the L superscript indicates the low-wage group (alternatively, white females and black males in this study).

Given (2) and (3), it is a simple matter to compute the portion of the differential explained by the regression: $\Sigma_j \beta_j^H \bar{X}_j^H - \Sigma_j \beta_j^L \bar{X}_j^L$ and the amount which is captured by the shift coefficient, $\beta_o^H - \beta_o^L$. The latter is typically attributed to discrimination (see, for example [14], ch. 5). However, we can carry our breakdown farther than this. The explained part of the differential comes from both differences in the coefficients, β_j^H and β_j^L, and differences in the average characteristics, $\bar{X}^H$ and $\bar{X}^L$. In particular:

$$\sum\nolimits_j \beta_j^H \bar{X}_j^H - \sum\nolimits_j \beta_j^L \bar{X}_j^L = \sum\nolimits_j \beta_j^H (\bar{X}_j^H - \bar{X}_j^L) + \sum\nolimits_j \bar{X}_j^L (\beta_j^H - \beta_j^L) \tag{4}$$

Here the first sum is the value of the advantage in endowments possessed by the high-wage group *as evaluated by the high-wage group's wage equation.*[3] The second sum is the difference between how the high-wage equation *would value* the characteristics of the low-wage group and how the low-wage equation *actually values* them. For brevity, I shall say that the first sum is "attributable to the endowments," while the second is "attributable to the coefficients." Note that the latter sum, which exists only because the

3 Any such breakdown of wage differentials runs into obvious index-number problems. For example, I could equally well have evaluated the differences in endowments by using the low-wage equation, i.e., by calculating

$$\Sigma_j \beta_j^L (\bar{X}_j^H - \bar{X}_j^L)$$

There is obviously no "right" and "wrong" answer here. My reason for adopting the alternative procedure is that the above formulation leaves as a residual:

$$(\bar{Y}^H - \bar{Y}^L) - \Sigma_j \beta_j^L (\bar{X}_j^H - \bar{X}_j^L)$$
$$= \beta_o^H - \beta_o^L + \Sigma_j \bar{X}_j^L (\beta_j^H - \beta_j^L) + \Sigma_j (\bar{X}_j^H - \bar{X}_j^L)(\beta_j^H - \beta_{\,j})$$

While we obviously could label the first two terms as "attributable to the coefficients," the interaction term has no obvious interpretation. The breakdown which I employ in the text, while in no sense theoretically "superior" to this, at least has the advantage that each term has a ready economic interpretation.

market evaluates differently the identical bundle of traits if possessed by members of different demographic groups, is a reflection of discrimination as much as the shift coefficient is.

In summary, the measures we shall use in Sections III and IV below are:

R = raw differential

$$= \beta_o^H + \Sigma_j \beta_j^H \bar{X}_j^H - (\beta_o^L + \Sigma_j \beta_j^L \bar{X}_j^L) = E + C + U$$

E = portion of differential attributable to differing endowments

$$= \Sigma_j \beta_j^H(\bar{X}_j^H - \bar{X}_j^L)$$

C = portion of differential attributable to differing coefficients

$$= \Sigma_j \bar{X}_j^L(\beta_j^H - \beta_j^L)$$

U = unexplained portion of the differential $= \beta_o^H - \beta_o^L$

D = portion of the differential attributable to discrimination $= C + U$

II. DATA AND ESTIMATION TECHNIQUES[4]

The data bank on which the present study is based is the Michigan Survey Research Center's "Panel Study of Income Dynamics."[5] This unique data source has several important advantages over most competing data sources. First, because income variables include many diverse sources of nonlabor income and because earnings variables confound differences in wage rates with differences in labor supply (which are themselves functions of wages), it is imperative that studies such as the present one use actual wage rates as the dependent variable.[6] Most survey data generated by the Bureau of the Census lack such a measure, which probably explains the paucity of empirical wage equations until very recently. The Survey of Economic Opportunity (SEO) has, however, provided the micro data for a number of recent regression studies of individual wages (see, for example, [9, 4, 11, 10]).

4 For more details on the precise definitions of the variables employed here, and a more thorough discussion of estimation problems, see the longer paper [1] upon which the present paper is based. It is available upon request.

5 For information on the survey techniques, collection, and processing of the data, see [13].

6 For example, use of earnings data instead of wage rates can seriously bias estimates of the real rates of return to education. Suppose, for example, that wages depend on education and labor supply depends on wages. Then, estimating the effect of education on *earnings* will over- or underestimate the impact on *wage rates* according as the labor supply curve is normal or backward-bending.

The Michigan data have only a minimal advantage over the SEO in this respect: the latter calculates the average wage earned in the week previous to the survey, while the former computes the average wage over the preceding year (1967). A more important advantage is that the Michigan data provide a rich set of variables pertaining to the individual's family background. These enable us to estimate a meaningful reduced-form equation which explains the wage rate only on the basis of characteristics which are truly exogenous to the individual (such as his father's education). The SEO is a much poorer source of family background data.[7] A third advantage of the Michigan data is its superior treatment of union membership—an important determinant of earnings. The SEO follows Census procedures in asking only *nongovernment* workers whether or not they are members of a union; thus government workers must be either eliminated or treated as nonmembers. Since government employees generally earn higher wages than private nonunion workers, the latter procedure would bias downward the estimated union-nonunion differential.[8] The Michigan survey asked the union membership question of every household head. Finally, the SRC tape offers data on the length of time each individual has worked for his present employer, thus giving us a direct measure of at least one part of the individual's total work experience.

Of course, there are some serious shortcomings of these data as compared, say, with the SEO. Most important is the small sample size, which makes the standard errors regrettably large in many cases. While the R^2s from my wage equations are quite comparable to those obtained with the SEO, the standard errors of individual coefficients are far larger. For this reason the inferences drawn in Sections III and IV are very tentative.[9] Finally, the Michigan tape lacks information on the industry of employment, a variable often deemed to be an important determinant of wages.[10]

To sharpen the estimates, I have selected a subsample from the SRC tape for use in my wage regressions. First, to eliminate household heads who may still be acquiring formal education, I dropped from the sample all households whose heads were younger than age 25. Also dropped from

7 Recent work by Samuel Bowles suggests that this may be a serious omission, but Bowles's work utilizes income as the dependent variable. See [5].

8 My thanks to Orley Ashenfelter for pointing this out to me. This may be one reason why the studies cited above [9, 4, 11, 10] all find smaller union-nonunion differentials than those reported here.

9 However, the procedures used to analyze the sex and race wage differentials automatically give lower weight to the coefficients with larger standard errors. So the analysis in Sections III and IV below may be more accurate than the regression coefficients given in the Appendix.

10 Like many other data banks, the SRC panel study also lacks an ability variable. However, some recent work by Gintis [7] and Griliches and Mason [8] suggests that this may not be an important omission.

consideration were household heads who were neither black nor white or who did not work for money in 1967.

I have argued above that the proper dependent variable in equations (2) and (3) is the hourly wage rate.[11] But what of the independent variables? This, of course, depends on one's (tacit or explicit) model of microeconomic wage determination. Most previous studies have been content to list certain obvious correlates of wages and estimate the resulting regression by ordinary least squares. But correlation does not imply causation, and if we are interested in the latter, we ought to specify a more precise model. The structural model which underlies the present study is the following:

$$\log w = f(Ed, Occ, J, M, V, T, Z) + u_1 \tag{5}$$
$$Ed = g(Occ, V, B, Z) + u_2 \tag{6}$$
$$Occ = h(Ed, J, V, B, Z) + u_3 \tag{7}$$
$$J = k(Occ, V, M, B, Z) + u_4 \tag{8}$$
$$M = l(Ed, Occ, V, B, Z) + u_5 \tag{9}$$
$$V = m(Ed, Occ, J, B, Z) + u_6 \tag{10}$$
$$T = n(Ed, Occ, J, M, V, Z) + u_7 \tag{11}$$

where w is the hourly wage; Ed is a vector of six educational dummy variables; Occ is a set of eight dummies for occupation; J is a dummy for vocational training; M is 1 for members of unions and 0 otherwise; V is 1 for veterans and 0 otherwise; T is a set of six dummies for tenure on the present job; B is a set of 13 family-background variables; Z is a set of other exogenous variables; and f, g, h, k, l, m, and n are all linear functions. In this model, w, Ed, Occ, J, M, V, and T are taken to be endogenous, while B and Z are exogenous. The elements of Z which enter the wage equation with nonzero coefficients are age, health, residence, and local labor market conditions.

While every reader can doubtless quarrel with some detail of this specification, the two most important features are relatively noncontroversial. Presumably, no one would wish to exclude from the structural wage equation any of the variables which I have included (though other investigators have added variables to the above list). The dichotomization of variables into endogenous and exogenous sets may require some comment, however, as it is rather novel in this context. In the intuitive model I have in mind, each individual is presented with endowments of human and nonhuman capitals and at some point in his life-cycle, jointly decides how far he wishes to pursue his formal education and to what occupational strata he aspires. Thus *Ed* and *Occ*, the two chief determinants of the wage rate, are endogenous and simultaneously determined. This seems very much in the

11 Like most other investigators, I have actually used the natural logarithm of the wage rate as the dependent variable.

spirit of the human-capital-theory approach to wage differences. Union membership is endogenous because each worker may opt for a "union job" at his discretion and because unions themselves may be selective about whom they admit.[12] Similar remarks hold for veteran status: some men volunteer, while others are selected by a not-quite-random process.

Once this simultaneous-equations viewpoint is adopted, the estimation problem seems quite straightforward: Ideally, equations (5) through (11) should be estimated as a simultaneous system by two-stage least squares or some similar technique. Unfortunately, this is not possible because, as the reader can verify, the wage equation is underidentified. The only variables omitted from (5) are the family-background variables, B. And this identification problem is by no means restricted to my particular model. Almost any specification of a micro wage equation will have difficulty excluding enough variables to identify the coefficients since these excluded variables (family background in my model) must be included in some other equation of the system. That is, they must be important determinants of education or occupation, etc., but have no direct impact on wage rates.

What can be done about this problem? In this study I have adopted two second-best procedures. The first is to estimate the reduced-form wage equation:

$$\log w = F(B, Z) + v_1 \tag{12}$$

instead of equation (5), by ordinary least squares. This is rigorously correct, and not without interest if we are concerned with the "ultimate causes" of wage differentials rather than the mechanisms through which these causes operate. Still, we would like to get a look inside the "black box," that is, to get some structural information about equation (5). To do this we need another, rather restrictive, assumption—that the error terms in equations (6) through (11) are all uncorrelated with u_1. As the reader may readily verify, if $E(u_1u_2) = E(u_1u_3) = E(u_1u_4) = E(u_1u_5) = E(u_1u_6) = E(u_1u_7) = 0$, then the system is block recursive and ordinary least squares is the best linear unbiased estimator of equation (5).

In summary, we present two wage-equation estimates. The reduced-form estimates—which are always unbiased—are obtained by ordinary least squares estimation of equation (12). The structural estimates—which are biased unless the error terms are uncorrelated—are obtained by ordinary least squares on equation (5). Very roughly speaking, the structural estimates can be thought of as the conditional expectation of (the log of) the wage, given the individual's present socioeconomic condition; and the reduced form can be thought of as the conditional expectation of (the log of) the wage, given the circumstances of his birth.

12 I have presented and discussed the results of estimating equation (9), the equation for union membership, in [3].

III. ANALYSIS OF THE WHITE-BLACK WAGE DIFFERENTIAL FOR MALES

This section breaks down the raw differential between white males and black males into its component parts, using the procedures outlined in Section I above.[13] The regression coefficients upon which the following tables are based are provided in the Appendix.

Table 1 decomposes the overall 50.5 percent wage advantage of whites in the manner implied by the structural estimates. One striking re-

13 One question not considered in Section I was the choice of the control group. For example, if a person can fall into one of eight mutually exclusive and exhaustive occupational categories, then we have eight possible ways to choose seven dummy variables to enter into the regression. My choices for the control group are all indicated in the Appendix. But a question arises as to whether a different choice of the control group would have altered my decompositions of the wage differentials.

While it is difficult to prove that this holds in an arbitrary regression using a set of K qualitative variables with n_k classes in each, one can show that in the simplest case of a *single dichotomous* classificatory variable, the choice of the control group does not matter. *Proof*: In this simple case, equations (2) and (3) become:

$$(2') \qquad Y_i^H = \beta_o^H + \beta_1^H X_{1i}^H + U_i^H$$

$$(3') \qquad Y_i^L = \beta_o^L + \beta_1^L X_{1i}^L + U_i^L$$

if group 2 is the control group. Under the alternative choice they would be:

$$(2'') \qquad Y_i^H = b_o^H + \beta_2^H X_{2i}^H + V_i^H$$

$$(3'') \qquad Y_i^L = b_o^L + \beta_2^L X_{2i}^L + V_i^L$$

My procedure is to divide the overall differential, $\overline{Y}^H - \overline{Y}^L$, into $E_1 = \hat{\beta}_1^H(\overline{X}_1^H - \overline{X}_1^L)$ and $\overline{Y}^H - \overline{Y}^L - E_1$. Using the other group as control, we would want instead $E_2 = \hat{\beta}_2^H(\overline{X}_2^H - \overline{X}_2^L)$. But since $X_{1i} + X_{2i} = 1$ for all i we have:

$$\overline{X}_2^H - \overline{X}_2^L = -(\overline{X}_1^H - \overline{X}_1^L)$$

And from the standard formula for the estimated slope coefficients, we can show:

$$\hat{\beta}_2^H = \hat{\beta}_1^H$$

So $E_2 = E_1$, and our dichotomization of the wage differential is the same for either choice of the control group.

TABLE 1
STRUCTURAL ANALYSIS OF THE WHITE-BLACK WAGE DIFFERENTIAL FOR MALES

Causal Factor		Amount Attributable	Amount Attributable to Endowments	Amount Attributable to Coefficients
Age		5.1%	−0.7%	5.8%
Region of residence		−3.6	5.1	−8.7
Education		19.8	14.5	5.3
Vocational training		2.1	0.2	1.8
Occupation		1.4	10.3	−8.9
Union membership		−5.7	−0.4	−5.3
Veteran status		−2.0	0.5	−2.5
Health		−0.6	−0.5	−0.1
Local labor market conditions		−0.7	−0.1	−0.6
Geographical mobility		0.9	0.5	0.5
Seasonal employment		−1.7	0.3	−2.0
Length of time on job		−36.2	0.7	−36.9
Subtotal		−21.1%	$E = +30.7\%$	$C = -51.5\%$
Shift coefficient	$U =$	+71.9%		
Total	$R =$	+50.8%	$D = C + U = 20.4\%$	

Notes: A + sign indicates advantage for whites; a − sign indicates advantage for blacks. Components may not add to totals due to rounding.

sult is that the regression (ignoring the constant) actually accounts for a 21.1 percent differential *in favor of blacks*. That is to say, if black males kept their current socioeconomic traits (including their less advantageous distributions of education and occupation), and kept the same wage equation as they now have but were given the shift coefficient of the whites, they would earn 21.1 percent more than white males. The table also shows that a 30.7 percent differential (about 60 percent of the total) can be attributed to blacks' inferior endowments, in the sense that black wages would average 30.7 percent less than white wages if the former were given the entire white wage equation. Closer examination reveals that the main reason why the regression calls for higher wages for blacks is that they gain more by having seniority in their present jobs. The regression coefficients in the Appendix further show that just about all of this difference is accounted for by the large gain in wages which accrues to black males from being in an occupation where tenure on the job is applicable (as opposed to the self-employed who were the control group). If this were ignored, the regression (ignoring the constant) would predict a 15.8 percent advantage for whites, a rather small difference.

Foremost among the factors contributing to the white male wage advantage is education, and the breakdown shows that most of this is due to the fact that whites have more education rather than to the higher rates of return on education which whites also receive. The other main advantage for white men is the more concave age-wage profile, a fact that has been noted in many studies of age-earnings profiles as well. Whites apparently gain more than blacks from experience in the labor force, and therefore they exhibit a more pronounced life-cycle pattern of wages.

Aside from tenure on the job, the most important factors favoring black males are union membership and geographical distribution across regions of the country. While blacks and whites are about equally likely to be union members, blacks gain much more from being in a union. The regression coefficients show an average union-nonunion differential of 39 percent for black males and 23 percent for white males. Both of these are larger than the differentials which have been estimated from the SEO (see [9, 4, 11, 10]). How much of this discrepancy is due to the superior data on union membership (see above), how much is due to omission of relevant variables fom the regression, and how much is due to the absence of interaction terms which allow the union-nonunion differential to vary across occupations[14] is an open question. However, the fact that Stafford's estimates [12], which used Survey Research Center data, are among the highest obtained to date lends some credence to the notion that differences in the conceptual basis of the data may well be the reason for the high differentials. As to the regional distribution, the fact that blacks are more heavily concentrated in the South (a disadvantage) is more than balanced by their higher coefficients for each region (as compared to the North Central region which is the control group).

Interestingly, the distribution of workers across occupations—at least at this relatively coarse level of disaggregation—contributed only negligibly to the white-black differential. This would appear to fly in the face of casual empirical observation, but in fact it does not. Decomposition of the meagre 1.4 percent occupational differential in favor of whites reveals that the superior white occupational distribution actually accounts for a substantial 10.3 percent wage advantage, most of which is neutralized by higher co-

14 Boskin [4], p. 470, suggests that ignoring the interactions between occupation and union membership (i.e., that union-nonunion differentials differ across occupations) results in a serious *overestimate* of the *average* union-nonunion differential. However, as Orley Ashenfelter points out to me, his reasoning is flawed. The figures in his Table 2 for union differentials in each occupation ignore the geographical area-union interactions which also appear in his regressions. That is, to his estimated union-nonunion differentials for each occupation in Table 2 one should add the average (over all regions) of the union differential in that occupation. That would (almost) always result in higher union differentials than he presents. Indeed, it is difficult to see how a weighted average of the union-nonunion differentials (weighted by occupation) could differ much from the overall average differential in the sample.

TABLE 2
REDUCED FORM ANALYSIS OF THE WHITE-BLACK WAGE DIFFERENTIAL FOR MALES

Causal Factor		Amount Attributable	Amount Attributable to Endowments	Amount Attributable to Coefficients
Age		83.6%	−1.1%	84.7%
Region of residence		−0.8	8.8	−9.6
Parents' income		5.6	0.5	5.1
Father's educaion		12.3	3.6	8.7
Place where grew up		16.2	−0.7	16.9
Number of siblings		−6.6	1.2	−7.7
Health		−0.9	−1.0	0.1
Local labor market conditions		−6.4	0.3	−6.6
Geographical mobility		6.8	2.9	3.9
Seasonal employment		−2.5	0.8	−3.4
Subtotal		+107.4%	$E = +15.2\%$	$C = +92.2\%$
Shift coefficient	$U =$	−56.7%		
Total	$R =$	+50.7%	$D = C + U = 35.5\%$	

Notes: A + sign indicates advantage for whites; a − sign indicates advantage for blacks. Components may not add to totals due to rounding.

efficients for occupational advancement for blacks.

Table 2 portrays a similar decomposition based on the reduced-form estimates. The results are strikingly different. When wages are explained by circumstances of birth and other exogenous characteristics, the shift coefficient (56.7 percent) now favors black men. The regression itself, apart from the constant, actually accounts for a 107.4 percent wage differential in favor of whites, and most of this comes from a superior wage equation rather than from superior endowments.

The dominant factor favoring whites is, of course, the age-wage profile. As white males mature from age 25 to age 46, wage rates rise steadily due to past investments in human capital and to experience. After that, wages decline with age as training becomes obsolete and physical conditions deteriorate. Blacks also show a concave age profile, but it is much less steep. Table 2 shows that these differing age profiles account for nearly 80 percent of the explained white-black wage differential, and almost all of the part attributable to the regression coefficients.

Not surprisingly, the other major factors favoring whites are almost all family-background indices: parents' income, father's educational attainment, and the place where the individual was brought up (whether on a farm, in a city, or in the South). What is surprising is that most of the advantage for whites comes from the coefficients rather than from the endowments. Whites' superior endowments of these three characteristics can explain only a 3.4 percent differential in their favor, while a 30.7 differential is explicable by the whites' more favorable coefficients. Blacks generally gain less by having an educated father, lose more by coming from a poor family, gain less by being born in an urban area, and so on. The only family-background variable favoring blacks is the number of siblings. While the sizes of black and white families are not very different, blacks actually reap gains from having brothers and sisters, while whites suffer wage reductions.

Note that since $D = 35.5$ percent and $R = 50.7$ percent in Table 2, the reduced form attributes about 70 percent (that is, .355/.507) of the raw white-black wage differential for men to the various dimensions of discrimination, including discrimination in education and occupation. By contrast, the structural equation, which measured only discrimination in wage rates, attributed only about 40 percent to discrimination ($D/R =$ .204/.508 in Table 1). Thus, we might tentatively break down the raw differential as follows: 30 percent to blacks' inferior endowments of exogenous variables (the ratio E/R in Table 2); 40 percent to outright discrimination in rates of pay (the ratio D/R in Table 1); and 30 percent to discrimination in achieving the other endogenous variables (such as education, occupation).

It is perhaps of some interest to compare these crude results to Duncan's attempt to break down the black-white *income* differential for adult males.[15] Using a somewhat different specification and different statistical methods, but a decomposition technique which is almost the same as mine, he attributed 26.6 percent of the differential to family-background variables,[16] 35.6 percent to divergent attainments of education and occupational status which could not be predicted by family background, and the remaining 37.8 pecent to discrimination in income. This seems remarkably comparable to my rough 30-30-40 percent breakdown,[17] especially in view of the many differences between the two studies.

IV. ANALYSIS OF THE MALE-FEMALE WAGE DIFFERENTIAL FOR WHITES

Table 3 utilizes the structural regressions to decompose the 45.6 percent raw wage differential for white males over white females. The nature of this differential is seen to be radically different from the structural race differ-

15 Duncan [6], pp. 84–110. Specifically, his sample was black versus *nonblack* males aged 25–64 years who were in the experienced labor force and whose fathers did not work on farms.

16 This groups "number of siblings" with the family-background variables, as I do. Duncan takes this to be an endogenous variable instead.

17 Of course, I have many exogenous variables aside from family background and more endogenous variables than just education and occupation.

TABLE 3

STRUCTURAL ANALYSIS OF THE MALE-FEMALE WAGE DIFFERENTIAL FOR WHITES

Causal Factor	Amount Attributable	Amount Attributable to Endowments	Amount Attributable to Coefficients
Age	87.5%	0.2%	87.3%
Region of residence	−0.9	0.2	−1.1
Education	21.3	−0.4	21.7
Vocational training	−3.5	−0.1	−3.4
Occupation	−4.1	5.1	−9.1
Union membership	3.3	3.9	−0.6
Veteran status	1.6	1.6	0
Health	2.0	0.0	1.9
Local labor market conditions	12.2	0.1	12.1
Geographical mobility	−0.9	0.5	−1.3
Seasonal employment	0.1	0.0	0.1
Length of time on job	−4.7	4.7	−9.4
Subtotal	113.8%	$E = 15.7\%$	$C = 98.1\%$
Shift coefficient	$U = -68.0\%$		
Total	$R = 45.8\%$	$D = C + U = 30.1\%$	

Notes: A + sign indicates an advantage for males; a − sign indicates an advantage for females. Numbers may not add to totals due to rounding.

ential. The independent variables account for a 113.8 percent differential in favor of males on the basis of "objective" characteristics, leaving a 68 percent advantage for white *females* to the unexplained shift coefficient. Breaking down the explained 113.8 percent further shows that only 15.7 percent is due to inferior endowments, while 98.1 percent comes from different coefficients in the two wage equations.

By far the largest part of the differential is accounted for by the age-wage profile. This is in marked contrast to the structural differential between the races, where age profiles were not terribly important. In our estimates, white women exhibit an almost flat (in fact, slightly convex) age-wage profile—that is, their wages do not show any tendency to rise over the life cycle, whereas wages for white men certainly do. Thus, the failure of women *in the same education-occupation category* to rise on the economic ladder over their working lives is seen to be the single largest cause of the male-female differential among whites.

There are two further factors which place men at a distinct advantage: education and conditions in the local labor market. In each case, the sexes barely differ in their endowments; it is the coefficients that matter. Men earn much larger wage increments for advancing to higher educational levels, and they are much less sensitive to local labor market conditions.[18]

Notably absent from this list of factors explaining the male-female differential is occupation. It has been widely held that the main way in which women are discriminated against in labor markets is not in rates of pay, but by being relegated to lower positions on the occupational ladder than their qualifications would merit (see, for example, [16, 15]). Yet Table 3 shows occupational factors as contributing a 4.1 percent wage advantage to white women. On closer examination, however, the apparent conflict disappears. It is indeed true that men have a superior distribution across occupations (accounting for a 5.1 percent wage advantage for men), but the regressions show that this is more than balanced by the greater gains reaped by women when they enter certain occupations (accounting for a 9.1 percent differential for women). Furthermore, part of the disparity in age-wage profiles mentioned above is surely a reflection of the failure of women to rise to the higher occupational strata *within* any of the broad occupational groupings.

Table 4 shows that endowments count for even less (essentially zero) and discrimination for even more when reduced-form estimates are applied to the sex differential. In fact, there are hardly any differences between the reduced form endowments of white men and white women.

Once again the age profile dominates all other considerations, and it now accounts for the entire wage difference between the sexes. Embodied in this reduced-form age-wage profile, of course, are any age-related differences in the attainment of the other endogenous variables, such as occupation, union membership, and job experience. Most of the family-background variables operate in favor of the females, and most of their impact comes through the coefficients. (Obviously, white women are born in about the same places to the same kinds of parents and in families of about the same size as are white men.) The regression coefficients show that white women gain more from having educated fathers and from being brought up in a city or in the South, and actually benefit (while men lose) from having poor parents and being raised on a farm. Other than age, the main advantage accruing to white men is their reduced sensitivity to local labor market conditions.

In summary, the regressions attribute none of the observed male-female wage differential for whites to differences in their endowments of exogenous variables (E is about 0 in Table 4), about two-thirds to outright discrimination in labor markets (the ratio D/R in Table 3), and about one-third of the differential to discrimination in attaining other endogenous variables such as occupational status and job seniority.

18 This last remark ought to be discounted heavily since our dummy variables are very imperfect indices of local labor market conditions.

TABLE 4

REDUCED FORM ANALYSIS OF MALE-FEMALE WAGE DIFFERENTIAL FOR WHITES

Causal Factor	Amount Attributable	Amount Attributable to Endowments	Amount Attributable to Coefficients
Age	132.5%	1.0%	131.5%
Region of residence	4.8	0.4	4.4
Parents' income	−5.4	−0.7	−4.7
Father's education	−9.5	−1.0	−8.5
Place of birth	−14.8	−2.6	−12.2
Number of siblings	3.4	−0.1	3.5
Health	0.6	−0.1	0.6
Local labor market conditions	15.5	0.0	15.5
Geographical mobility	4.1	2.8	1.3
Seasonal employment	0.6	0.1	0.6
Subtotal	131.7%	$E = -0.3\%$	$C = 132.0\%$
Shift coefficient	$U = 86.2\%$		
Total	$R = 45.4\%$	$D = C + U = 45.8\%$	

Notes: A + sign indicates an advantage for males; a − sign indicates an advantage for females. Numbers may not add to totals due to rounding.

REFERENCES

1. A. S. Blinder. "Estimating a Micro Wage Equation: Pitfalls and Some Provisional Estimates." Princeton University Econometric Research Program, Research Memorandum No. 131, November 1971.
2. ———. *Towards an Economic Theory of Income Distribution*. Cambridge, Mass.: MIT Press, forthcoming 1974.
3. ———. "Who Joins Unions?" Princeton University Industrial Relations Section, Working Paper No. 36, February 1972.
4. M. J. Boskin. "Unions and Relative Real Wages." *American Economic Review* 62 (June 1972): 466–72.
5. S. Bowles. "Schooling and Inequality from Generation to Generation." *Journal of Political Economy* 80:2 (May-June 1972): S219–51.
6. O. D. Duncan. "Inheritance of Poverty or Inheritance of Race?" In *On Understanding Poverty*, ed. D. P. Moynihan. New York: Basic Books, 1968.
7. H. Gintis. "Education, Technology and the Characteristics of Worker Productivity." *American Economic Review* 61 (May 1971): 269–79.
8. Z. Griliches and W. M. Mason. "Education, Income and Ability." *Journal of Political Economy* 80:2 (May-June 1972): S74–103.
9. R. E. Hall. "Wages, Income and Hours of Work in the U.S. Labor Force." In *Labor Supply and Income Maintenance*, eds. G. Cain and H. Watts. Chicago: Markham, 1973.

10. M. D. Hurd. "Changes in Wage Rates Between 1959 and 1967." *Review of Economics and Statistics* (May 1971): 189–99.
11. R. Oaxaca. "Sex Discrimination in Wages." In *Discrimination in Labor Markets,* eds. O. Ashenfelter and A. Rees. Princeton, N.J.: Princeton University Press, forthcoming 1973.
12. F. Stafford. "Concentration and Labor Earnings: Comment." *American Economic Review* 58 (March 1968): 174–80.
13. Survey Research Center, University of Michigan. *A Panel Study of Income Dynamics: Study Design, Procedures, Available Data: 1968–1970 Interviewing Years (Waves I–III).* Ann Arbor: The Center, 1970.
14. L. C. Thurow. *Poverty and Discrimination.* Washington: Brookings Institution, 1969.
15. F. B. Weisskoff. " 'Women's Place' in the Labor Market." *American Economic Review* 62 (May 1972): 161–66.
16. H. Zellner. "Discrimination Against Women, Occupational Segregation and the Relative Wage." *American Economic Review* 62 (May 1972): 157–160.

APPENDIX

Tables A-1 and A-2 list, respectively, the structural and reduced-form wage regressions upon which the analysis in Sections III and IV is based. Coefficients are presented only for white males, black males, and white females. Results from estimating wage equations for black females and for the pooled sample, with race and sex entering as dummy variables only, are reported in the expanded version of this paper.[19]

19 See fn. 4 in the text.

TABLE A-1

STRUCTURAL COEFFICIENTS (Standard Errors)
(Dependent Variable: Natural Log of Wage)

Independent Variable	White Males	Black Males	White Females
Constant	4.86	4.14	5.54
	(.20)	(.45)	(.64)
Age	.0271	.0296	−.0110
	(.0078)	(.0195)	(.0240)
$(\text{Age})^2$	−.00029	−.00038	.00009
	(.00008)	(.00022)	(.00026)
Region			
Northeast	.061	.144	.109
	(.038)	(.107)	(.108)
North Central	0	0	0

TABLE A-1 (*Continued*)

Independent Variable	White Males	Black Males	White Females
South	−.107	−.013	−.090
	(.038)	(.088)	(.106)
West	−.011	.189	−.006
	(.042)	(.117)	(.113)
Local labor market			
Low unemployment	−.024	.014	−.246
	(.031)	(.054)	(.098)
High unemployment	−.095	−.168	−.255
	(.051)	(.097)	(.151)
Low wage	.010	−.207	.231
	(.041)	(.068)	(.122)
High wage	.073	.315	.203
	(.046)	(.109)	(.136)
Geographical mobility			
Moved for job	.025	−.056	.032
	(.030)	(.070)	(.100)
Refused to move for job	.016	.088	.230
	(.043)	(.083)	(.162)
Health impediments			
Obvious disfigurement	−.119	.004	−.183
	(.053)	(.103)	(.131)
Some work limitation	−.081	−.227	−.189
	(.038)	(.082)	(.135)
Sick recently	−.013	.049	−.030
	(.050)	(.074)	(.140)
Education			
0–5 years	0	0	0
6–8 years	.075	−.015	.179
	(.065)	(.065)	(.227)
9–11 years	.133	.175	.247
	(.043)	(.063)	(.128)
12 years	.084	.110	.022
	(.040)	(.075)	(.109)
Some college	.155	.012	.187
	(.042)	(.115)	(.119)
College graduate	.212	.136	−.109
	(.059)	(.259)	(.180)
Advanced degree	.164	.248	.241
	(.070)	(.323)	(.218)
Occupation			
Professional, technical	.049	.159	.130
	(.053)	(.174)	(.332)

TABLE A-1 (*Continued*)

Independent Variable	White Males	Black Males	White Females
Managers, officials, proprietors	.124	.567	.133
	(.049)	(.344)	(.354)
Self-employed businessmen	–.395	–.019	–.290
	(.069)	(.214)	(.412)
Clerical, sales	–.085	.011	–.024
	(.049)	(.115)	(.305)
Craftsmen	0	0	0
Operatives	–.162	–.193	.014
	(.040)	(.067)	(.309)
Laborers, service, farmhands	–.333	–.123	–.263
	(.053)	(.067)	(.305)
Farm managers	–.719	–1.051	—
	(.076)	(.213)	
Union member	.228	.393	.276
	(.032)	(.056)	(.122)
Veteran	.029	.099	—
	(.028)	(.053)	
Seasonal employment	–.050	.050	–.055
	(.038)	(.057)	(.107)
Vocational training	.031	–.069	.147
	(.032)	(.061)	(.085)
Tenure on the job			
Self-employed	0	0	0
Less than 6 months	–.085	.393	–.027
	(.065)	(.139)	(.237)
6–18 months	.025	–.103	.047
	(.069)	(.101)	(.146)
10–42 months	.020	.122	–.075
	(.060)	(.089)	(.129)
4–9 years	.056	.020	.189
	(.045)	(.073)	(.119)
10–20 years	.078	–.001	.024
	(.041)	(.068)	(.110)
Over 20 years	.081	.082	.125
	(.044)	(.081)	(.151)
R^2	.498	.467	.449
s.e.	.437	.466	.465
N	1239	467	194

Note: The coefficients for education and tenure-on-the-job variables are *increments* from advancing from one category to the next higher one.

TABLE A-2
REDUCED FORM COEFFICIENTS (Standard Errors)
(Dependent Variable: Natural Log of Wage)

Variable	White Males	Black Males	White Females
Constant	4.75	5.32	5.62
	(.22)	(.45)	(.59)
Age	.0427	.0110	–.0160
	(.0093)	(.0200)	(.0250)
$(\text{Age})^2$	–.00047	–.0002	.00014
	(.00010)	(.0002)	(.00027)
Region			
Northeast	.080	.130	–.007
	(.050)	(.119)	(.117)
North Central	0	0	0
South	–.205	–.099	–.235
	(.065)	(.101)	(.158)
West	–.024	.198	–.104
	(.053)	(.124)	(.120)
Local labor market			
Low unemployment	.015	.150	–.249
	(.040)	(.060)	(.102)
High unemployment	–.087	–.082	–.319
	(.060)	(.104)	(.159)
Low wage	.018	–.226	.369
	(.050)	(.077)	(.132)
High wage	.159	.400	.165
	(.057)	(.118)	(.141)
Geographical mobility			
Moved for job	.144	–.109	.034
	(.036)	(.074)	(.105)
Refused to move for job	.134	.059	.218
	(.053)	(.088)	(.167)
Health impediments			
Obvious disfigurement	–.201	–.018	–.189
	(.066)	(.111)	(.131)
Some work limitation	–.176	–.216	–.231
	(.048)	(.090)	(.138)
Sick recently	–.028	–.018	–.053
	(.062)	(.080)	(.149)
Seasonal employment	–.136	.032	–.177
	(.046)	(.063)	(.110)
Siblings			
Number if $\leqq 7$	–.003	.013	–.012
	(.009)	(.013)	(.023)

TABLE A-2 (*Continued*)

Variable	White Males	Black Males	White Females
Eight or more	–.071	.050	–.153
	(.059)	(.074)	(.157)
Father's education			
0–5 years	0	0	0
6–8 years	.090	–.015	.166
	(.055)	(.061)	(.153)
9–11 years	.163	–.055	.161
	(.067)	(.125)	(.160)
12 years	–.015	.109	–.012
	(.080)	(.171)	(.186)
Some college	–.143	.303	.178
	(.095)	(.392)	(.244)
College graduate	.228	.016	.282
	(.125)	(.410)	(.276)
Advanced degree	.082	—	–.874
	(.176)		(.356)
Parents' economic status			
Poor	–.012	–.075	.136
	(.036)	(.075)	(.092)
Rich	.060	.033	.041
	(.054)	(.107)	(.111)
Place where grew up			
In the South	.045	–.177	.070
	(.059)	(.090)	(.155)
On a farm	–.214	–.112	.063
	(.040)	(.064)	(.107)
In a city	.099	–.017	.222
	(.040)	(.060)	(.093)
R^2	.200	.329	.335
s.e.	.550	.516	.500
N	1239	467	194

Note: Coefficients for father's education indicate *incremental* impact of advancing to next highest educational category.