DOCUMENT RESUME

ED 132 432 CE 009 268

AUTHOR	Brown, Randall S.; And Others
TITLE	Measuring Wage and Occupational Discrimination: A Comprehensive Approach. Institute for Research on Poverty Discussion Papers.
INSTITUTION	Wisconsin Univ., Madison. Inst. for Research on Poverty.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C.
PUB DATE	Aug 76
NOTE	30p.
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Employment Level; *Employment Patterns; Employment Practices; Employment Qualifications; *Equal Opportunities (Jobs); Females; Labor Problems; Males; *Mathematical Models; Measurement; *Salary Differentials; *Sex Discrimination; Wages; Working Women

ABSTRACT

Many economists have tried to explain existing wage differentials between men and women. A new approach compares the relative importance of occupational discrimination with that of wage discrimination. This model allows for variation both in occupational distribution and in wages resulting from differences in job qualifications and productivity indicators. It was demonstrated that the usual approach to wage discrimination is a special case of this general model with some restrictive implicit assumptions. A multinomial logit model was used to stimulate the occupational distribution of women that would exist if they faced the same structure of occupational determination as men. Results indicate that there would be more women in managerial and skilled labor jobs and fewer women clerical and service workers. Wages were then estimated as a function of productivity measures for men and women in each occupation so that the components of the wage differential could be calculated. Results indicate that almost the entire differential could be eliminated by ending both forms of discrimination, with occupational discrimination accounting for one-third to one-half of the differential and pure wage discrimination the remainder. (WL)

Measuring Wage and Occupational Discrimination:

A Comprehensive Approach

Randall S. Brown
Marilyn Moon
Barbara S. Zoloth

August, 1976

This paper was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin-Madison by the Department of Health, Education, and Welfare pursuant to the Economic Opportunity Act of 1964. The authors are at Mathematica Policy Research, Princeton; University of Wisconsin-Milwaukee; and University of California at Davis, respectively. They wish to acknowledge Nancy Williamson for her programming assistance and Douglas Gomery for comments on an earlier draft.

ABSTRACT

Differences in the occupational distributions of men and women represent an important source of male-female wage differentials. Although some studies of sex discrimination in the labor market deal explicitly with occupational segregation and some deal directly with the problem of unequal pay for equal work, few have concerned themselves with both issues simultaneously. This study develops an approach that decomposes the aggregate male-female differential such that the effects of occupational barriers are distinguishable from the effects of wage discrimination. Our model allows for variation both in occupational distribution and in wages resulting from differences in job qualifications and productivity indicators. The usual approach to explaining wage differentials is shown to be a special case.

Our approach uses multinomial logit analysis to predict occupational attainment from a set of personal characteristics. This estimation is performed on the National Longitudinal Survey (NLS) sample of mature men and the resulting parameters are used to simulate the occupational distribution of the NLS mature women that would exist if these women faced the same structure of occupational determination as the men. Wages are then estimated as a function of productivity measures for both men and women within each occupation so that the components of the wage differential can be calculated. We find that almost the entire differential would be eliminated by ending both forms of discrimination, with occupational discrimination accounting for one-third to one-half of the differential and pure wage discrimination the remainder.

MEASURING WAGE AND OCCUPATIONAL DISCRIMINATION:

A COMPREHENSIVE APPROACH

In recent years economists have attempted to explain the well-known discrepancy between average wages of men and women. Much of the research in this area concentrates on the issue of unequal pay for equal work. However, differences in the occupational distributions of men and women also comprise an important source of aggregate wage differentials. Since most studies emphasize only one of these areas, economists have been unable to measure the relative size of each effectively or to formulate a comprehensive indicator of discrimination after accounting for "justifiable" wage and occupation differences. This paper develops such a comprehensive approach--one that decomposes the male-female differential such that the effects of occupational barriers are distinguishable from the effects of wage discrimination. Our model incorporates the usual approach to wage discrimination as a special case and refines the typical approach to measuring occupational discrimination, making use of a model that predicts occupational attainment.

Earlier discrimination studies can be subdivided into two broad categories. The first, the wage approach, attempts to separate the wage differential into two portions: productivity differences and discrimination. Studies of this type include Blinder (1973), Sawhill (1973), Malkiel and Malkiel (1973) and Fuchs (1971).[1] The basic method is to regress wages on personal characteristics for one or both sexes and to use the estimated coefficients to determine the portion attributable to pure wage discrimination. This wage approach fails to provide a full accounting of the justifiable portion of the wage differential because it employs inadequate procedures (or none at all) to account for differences in the occupational distribution of men and women. The second, the distribution approach, addresses these occupational differences directly. Gwartney and Stroup (1973), Zellner (1972) and Sanborn (1964) multiply the mean wage for each occupation for women by the proportion of the male labor force in that occupation. These products are then summed over occupations to arrive at the mean women's wage that would exist if women had the same occupational distribution as men. Although these studies do consider the effects of occupational distributions on wage differentials, they make only ad hoc adjustments for differences in average productivity traits in men and women.

One of the stumbling blocks to studies on occupational discrimination has been the lack of empirical work in the area of occupational attainment. Boskin (1974) sets up a conditional logit model to predict occupational choice, a somewhat different problem from occupational attainment. In a study which is more relevant for our purposes, Schmidt and Strauss (1975) estimate a multinomial logit model of occupational attainment. It is, however, limited to four explanatory variables and five occupational categories.

In Section I of this paper, we suggest a more general method that incorporates the desirable features of both the discrimination approaches described above and corrects for some of their shortcomings. We use a model proposed by Blinder (1973) as the starting point for our analysis. Section II develops a model to predict occupational attainment from a set of personal characteristics. The approach is similar to that of Schmidt and Strauss (1975) but with more occupational categories and explanatory variables. The parameters of the model a estimated using the sample of mature men from the National Longitudinal Survey (NLS). The resulting parameter estimates are then used to simulate the occupational distribution of women that would exist if women faced the same employment possibilities as men, ignoring the demand side of the market and differences in tastes. After calculating earnings functions in Section III to measure wage discrimination, Section IV combines these wage and occupational estimates into the general model. The results enable a comparison of the relative importance of each of these in explaining wage differentials.

I. A General Model

To facilitate the presentation and the comparison of our model to earlier findings, we refer to a study by Blinder (1973) as representative of others which examine earnings differentials. The usual method of measuring wage discrimination is to directly estimate a model such as

$$w^M = \alpha^M + X^{M'}\beta^M + \varepsilon^M$$
$$w^F = \alpha^F + X^{F'}\beta^F + \varepsilon^F \qquad (1)$$

where w^t is an $N^t \times 1$ vector of wages and X^t is a $K \times N^t$ set of personal characteristics, and $t = M, F$. Superscripts M and F refer to men and women, respectively. Using estimates of the coefficients a^F, a^M, b^F and b^M, one can compute the portion of the overall mean wage differential which is explained by the regression, $\bar{x}^M b^M - \bar{x}^F b^F$ where $\bar{x}^M$ and $\bar{x}^F$ are $1 \times K$ row vectors of variable means. The amount due to the difference in intercepts is $a^M - a^F$. Blinder states that the latter difference is typically attributed to discrimination, but that this breakdown can be extended to differences in the other coefficients as follows:

$$\bar{x}^M b^M - \bar{x}^F b^F = (\bar{x}^M - \bar{x}^F)b^M + \bar{x}^F(b^M - b^F).$$

Blinder then writes the raw mean wage differential, R, as:

$$\underset{(R)}{\bar{w}^M - \bar{w}^F} = \underset{(U)}{(a^M - a^F)} + \underset{(E)}{(\bar{x}^M - \bar{x}^F)b^M} + \underset{(C)}{\bar{x}^F(b^M - b^F)}, \qquad (2)$$

where E is the portion of the differential due to endowments, C is the portion attributable to differing coefficients, and U is the unexplained portion. Blinder defines $D = C + U$ as his measure of the portion of the total differential due to discrimination.

However, by explicitly including occupational differences, we can rewrite the total differential as:

$$
\begin{aligned}
\bar{w}^M - \bar{w}^F &= \sum_j (p_j^M \bar{w}_j^M - p_j^F \bar{w}_j^F) \\
&= \sum_j (p_j^M a_j^M - p_j^F a_j^F) + \sum_j (p_j^M \bar{x}_j^M b_j^M - p_j^F \bar{x}_j^F b_j^F) \\
&= \underbrace{\sum_j p_j^F (a_j^M - a_j^F)}_{(\mathrm{I})} + \underbrace{\sum_j p_j^F (\bar{x}_j^M b_j^M - \bar{x}_j^F b_j^F)}_{(\mathrm{WAGES})} \qquad (3) \\
&\quad + \underbrace{\sum_j (\bar{x}_j^M b_j^M + a_j^M)(p_j^M - p_j^F)}_{(\mathrm{OCC})} \\
&= \mathrm{I} + \underbrace{\sum_j p_j^F \bar{x}_j^F (b_j^M - b_j^F)}_{(\mathrm{WD})} + \underbrace{\sum_j p_j^F (\bar{x}_j^M - \bar{x}_j^F) b_j^M}_{(\mathrm{PD})} + \mathrm{OCC}
\end{aligned}
$$

where p_j^M and p_j^F are the sample proportions of men and women in the j occupation, $\bar{x}_j^M$ and $\bar{x}_j^F$ are vectors of sample means of the independent variables for men and women in the j occupation, and $\bar{w}_j^M$ and $\bar{w}_j^F$ are the mean wages received. Finally, a_j^M, b_j^M, a_j^F and b_j^F are ordinary least squares estimates of the coefficients in (1) for the j occupation.

Variables $\bar{w}^M$ and $\bar{w}^F$ are the grand mean wages paid to all men and women in the sample. The interpretation of the components of (3) is as follows:

WAGES = WD + PD = portion of wage differential due solely to explained differences in wages

OCC = portion of wage differential due to differences in occupational distribution

I = unexplained portion of wage differential (differences in intercepts)

I + WD = portion due to wage discrimination

PD = portion due to productivity differences between men and women

This breakdown now gives us an approach for measuring wage discrimination that takes into account differences in occupational distribution.

We can, in fact, write Blinder's model as a special case of (3). If we assume

$$\beta_j^M = \beta^M,\ \beta_j^F = \beta^F,\ \alpha_j^M = \alpha^M,\ \alpha_j^F = \alpha^F,$$

and either

$$(p_j^M = p_j^F) \text{ or } (\bar{x}_j^M = \bar{x}^M \text{ and } \bar{x}_j^F = \bar{x}^F) \text{ for } j = 1, \ldots, J$$

then

$$OCC = 0,\ I = a^M - a^F,\ WD = \bar{x}^F(b^M - b^F),\ \text{and } PD = (\bar{x}^M - \bar{x}^F)b^M.$$

With these substitutions (3) reduces exactly to (2), and we have Blinder's model. Thus we see that Blinder is assuming that the same relationship between wages and personal characteristics holds for all occupations and that either no differences in occupational distribution exist or that for each sex the mean values of the independent variables are identical for all occupations. Both assumptions seem quite untenable.

The model in (3), a mixture of the two usual approaches taken to explain the wage differential, is thus more appealing than either separate approach. Moreover, the model can be further decomposed. In a manner similar to the above disaggregation, we are able to allocate the share of wage differentials resulting from occupational segregation into the portion due to discrepancies in qualifications between men and women and the portion attributable to occupational discrimination. That is, we can write:

$$OCC = \sum_j \bar{w}_j^M (p_j^M - p_j^F) = \underbrace{\sum_j \bar{w}_j^M (p_j^M - \hat{p}_j^F)}_{(QD)} + \underbrace{\sum_j \bar{w}_j^M (\hat{p}_j^F - p_j^F)}_{(OD)}$$

where $\hat{p}_j^F$ is the proportion of women in the sample who would be in occupation j if women faced the same occupational attainment structure as men. The components of OCC can thus be described as follows:

QD = portion due to differences in qualifications for occupation

OD = portion due to occupational discrimination.

Consequently, the final decomposition is:

$$\bar{w}^M - \bar{w}^F = \sum_j p_j^F(a_j^M - a_j^F) + \sum_j p_j^F \bar{x}_j^F(b_j^M - b_j^F) + \sum_j p_j^F(\bar{x}_j^M - \bar{x}_j^F)b_j^M$$

$$\text{(I)} \qquad \text{(WD)} \qquad \text{(PD)} \qquad (4)$$

$$+ \sum_j \bar{w}_j^M(p_j^M - \hat{p}_j^F) + \sum_j \bar{w}_j^M(\hat{p}_j^F - p_j^F).$$

$$\text{(QD)} \qquad \text{(OD)}$$

However, the $\hat{p}_j^F$ are unobserved, and must be estimated. The approach we use first sets up a model to predict occupational attainment for men on the basis of a number of personal characteristics. We then predict the occupational distribution of women, assuming that they are subject to the same structure of determination as men, i.e. using the parameters estimated for the male sample. This approach is described in detail below.

II. Predicting Occupational Distribution

The Multinomial Logit Model

A regression model is inappropriate to predict occupations since this dependent variable is strictly qualitative (i.e., neither cardinal nor ordinal). We use, instead, a multinomial logit model, which allows us to predict the set of J probabilities, p_{ij}, such that individual i, with a vector of K personal characteristics, x_i, will work in each of the J possible occupations. Although our model is generally applicable to both men and women, we estimate parameters for men only and therefore drop the use of superscripts in this section, "M" being implicit.

Nerlove and Press (1973, p. 19) define the probability function as:

$$p_{ij} = \text{prob}(y_i = a_j) = \frac{e^{x_i'\gamma_j}}{\sum_{k=1}^{J} e^{x_i'\gamma_k}} \qquad i = 1, \ldots, N; \quad j = 1, \ldots, J \tag{5}$$

where i refers to the i observation, a_j is the j possible value of the dependent variable, y, N is the number of observations, and γ_k is a K x 1 vector of coefficients. The dependent variable y can take on any of the J possible values $a_1, \ldots, a_J$, each corresponding to a different occupation. This model, together with the normalization $\sum_j \gamma_j = 0$ necessary for identification, can be estimated by maximum likelihood methods and used for the purposes described above.[2] The likelihood function corresponding to (5) is:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} p_{ij}^{v_{ij}}, \text{ where } v_{ij} = \begin{cases} 1 \text{ if } y_i = a_j \\ 0 \text{ otherwise} \end{cases}$$

The first order conditions result from maximization of log L subject to the normalization.[3]

The Data and the Empirical Specification

Data limitations have consistently plagued studies of sex discrimination, especially because labor market histories of women are rarely available.[4] Fortunately, the National Longitudinal Survey panel studies (NLS) provides data on a number of relevant variables not available in other surveys. We have constructed the following independent variables from information on the NLS files for the samples of mature men and women:

KIDS = number of children

OCFATH = Duncan index of father's occupation when individual was age 15

YRSSCH = highest grade completed

VOCTRG = number of months of vocational training

RES15 = dummy variable for residence at age 15: urban = 1, rural = 0

IESUM = sum of eleven Rotter scale variables pertaining to attitude[5]

IEPERS = sum of four Rotter scale variables pertaining specifically to motivation

EXPER = labor market experience prior to current job, defined as age at which current job was taken minus YRSSCH minus VOCTRG/12 minus 6

Our sample contains 2277 white males aged 45 through 59. Although a constant term is not necessary, we include one for two reasons. First, summing across individuals the predicted probabilities of being in any given occupation produces the sample frequency for that occupation.[6] Second, the presence of the constant term allows the predicted probabilities that individuals are in two different occupations to vary even if the estimated coefficients are identical for all explanatory variables (or the individual has $x_i = 0$). This difference could be due either to missing variables or to differences in aggregate demand for the two types of jobs.

We specify the number of possible occupations by condensing the twelve major census occupational categories into eight (due to a scarcity of observations in four of the twelve categories) as follows, where the numbers in parentheses indicate the range of three-digit occupational codes:[7]

1. Professional, Technical and Kindred Workers (0-195)
2. Managers and Proprietors (200-245)
3. Clerical and Kindred Workers (301-375)
4. Salespersons (380-395)
5. Craftsmen, Foremen, Kindred Workers (401-595)
6. Operatives and Kindred Workers (600-795)
7. Service Workers (801-895)
8. Laborers (901-994)

The Estimation Results

Table 1 presents the estimates of the coefficients, their t-ratios and ranking of the coefficients by occupation. The larger the variable, the more likely it is that an individual is in the occupation with the higher coefficient. Thus, for exa there is a monotonic decrease in the coefficients on father's occupation indicating that the higher the socioeconomic status of his father's occupation, the more likely an individual is to be in a "prestigious" occupation. Other findings include the expected result that if an individual grows up in an urban area he is less likely to be in farming than any other occupation. Most of the other coefficients on RES15 are not significantly different from 0, indicating that growing up in an urban rather than a rural area does not lead people to choose one occupation over another, except for occupation 4. Other things being equal, it is more likely that an individual will be a salesperson than anything else if he grew up in an urban area.

The more years of schooling and vocational training an individual has, the greater is the likelihood that he will be in the professional category. For craftsmen training is very important, but schooling is not. The results for the experience variable are more difficult to interpret. It appears that the old end up in menial jobs while the young get the managerial and professional positions. However, our experience variable includes only years in the labor force prior to obtaining the current job. Individuals in professional positions often have acquired them early in their career, since formal education rather than experience is the primary requirement. Conversely, individuals in menial positions may be unlikely to remain in them, thus leading to frequent job changes and relatively high average values for EXPER. The rankings for motivation (IEPERS) show that the more highly motivated an individual is, the higher the relative likelihood that he is in a managerial position. The attitude (IESUM) variable does not give results that are readily interpreted.

TABLE 1

Maximum Likelihood Estimates of Coefficients in the Multinomial Logit Model
(t-ratios in parentheses)

Occupation	CONST	KIDS	OCFATH	YRSSCH	VOCTRG	RES15	IESUM	IEPERS	EXPER
1	-6.30* (-10.00)	-.001 (-0.02)	.0095* (2.82)	.534* (15.50)	.0241* (4.62)	-.072 (-0.42)	.0012 (0.05)	-.088 (-1.57)	-.011 (-1.43)
2	2.27* (7.04)	.058* (2.40)	.0053* (2.31)	-.013 (-0.68)	-.0028 (-0.69)	-.485* (-4.58)	-.0290* (-2.19)	.043 (1.45)	-.034* (-7.86)
3	-1.01 (-1.68)	-.057 (-1.14)	.0002 (-0.06)	.086* (2.52)	.0024 (0.36)	.276 (1.54)	-.0011 (-0.05)	.038 (0.70)	-.026* (-3.27)
4	-1.38* (-2.16)	-.057 (-1.09)	-.0003 (-0.08)	.105* (2.94)	.0017 (0.24)	.461* (2.49)	-.0023 (-0.09)	-.069 (-1.17)	.017* (1.98)
5	3.19* (9.90)	.011 (0.46)	-.0009 (-0.39)	-.148* (-7.88)	.0101* (2.89)	.071 (0.70)	-.0051 (-0.39)	-.022 (-0.74)	-.008 (-1.91)
6	2.33* (6.46)	.080* (3.16)	-.0018 (-0.64)	-.199* (-9.45)	-.0146* (-2.66)	-.002 (-0.01)	.0170 (1.17)	.022 (0.68)	-.001 (-0.24)
7	-0.14 (-0.22)	-.088 (-1.80)	-.0029 (-0.63)	-.082* (-2.30)	-.0061 (-0.74)	.340 (1.81)	-.0183 (-0.74)	.043 (0.77)	.039* (4.56)
8	1.04 (1.86)	.054 (1.49)	-.0090 (-1.62)	-.283* (-8.77)	-.0150 (-1.62)	-.590* (-2.62)	.0375 (1.65)	.034 (0.68)	.025* (3.27)

Occupational Rankings
by Size of Coefficient on:

OCFATH	YRSSCH	VOCTRG	EXPER	IESUM	IEPERS
1	1	1	7	8	2
2	4	5	8	6	7
3	3	3	4	1	3
4	2	4	6	3	8
5	7	2	5	4	6
6	5	7	1	5	5
7	6	6	3	7	4
8	8	8	2	2	1

*Denotes coefficient significant at the 5 percent level

The Predictions

With these coefficients we can now predict the occupational distribution of women, although a decision must first be made as to how to allocate individuals to occupations. One option is to allocate each individual to that occupation for which the predicted probability is highest. However, this approach correctly predicts occupations for only 39.4 percent of the men and fails to assign any individuals to occupations 3, 4 and 7.[8] A better alternative for predicting the occupational distribution is to sum the predicted probabilities of being in each occupation over all individuals. The first order conditions require that this sum equal the actual number of individuals in that occupation. Since our objective is to predict the occupational distribution of women using the model estimated on the men's data, we want to use that model which performs best in predicting the men's distribution. The implication is that, if women faced the same occupational structure as men, we would expect a sample of women to roughly satisfy this first order condition. Thus, the discrepancy between the actual and predicted occupational distributions for women can be attributed to differences between men and women in accessibility to different occupations, if occupational tastes are identical for working men and women.[9]

The alternative predicted occupational distributions for women are displayed in Table 2.[10] Table 3 presents the same results for the men. Line 1 of each table indicates the actual distributions. Predicted distributions using the logistic model of occupational attainment are given for the sum of probabilities method (line 2) and the maximum probability method (line 3). Finally, to see how sensitive our estimates of occupational distribution are to the logistic specification, results from a discriminant analysis calculation are displayed in line 4 of Tables 2 and 3.[11] Discriminant analysis does a somewhat better job (in terms of predicted distribution) than the logit model to which it most closely corresponds (line 3), but the results are very similar in pattern, and both

TABLE 2

Women's Predicted Occupational Distributions

[Proportion of Sample Size in Parentheses]

	Allocation Method	Occupation 1	2	3	4	5	6	7	8
1.	Actual Distribution	292 (.12)	121 (.05)	986 (.40)	175 (.07)	29 (.01)	385 (.16)	369 (.15)	78 (.03)
2.	$\sum_i \hat{p}_{ij}$, Logit Estimates	223 (.09)	800 (.33)	135 (.06)	106 (.04)	556 (.23)	437 (.18)	80 (.03)	97 (.04)
3.	$\max_j(p_{ij})$, Logit Estimates	196 (.08)	1549 (.64)	0 ---	0 ---	423 (.17)	266 (.11)	0 ---	1 ---
4.	$\max_j(p_{ij})$, Discriminant Analysis	176 (.07)	1330 (.55)	5 ---	2 ---	429 (.18)	428 (.18)	1 ---	64 (.03)

TABLE 3

Men's Predicted Occupational Distributions

[Proportion of Sample Size in Parentheses]

	Allocation Method	Occupation							
		1	2	3	4	5	6	7	8
1.	Actual Distribution	208 (.09)	581 (.26)	117 (.05)	108 (.05)	583 (.26)	422 (.19)	113 (.05)	145 (.06)
2.	$\sum_i \hat{p}_{ij}$, Unweighted Logit Estimates (Same as line 1)	208 (.09)	581 (.26)	117 (.05)	108 (.05)	583 (.26)	422 (.19)	113 (.05)	145 (.06)
3.	$\max_j(p_{ij})$, Unweighted Logit Estimates	233 (.10)	839 (.37)	0 ---	0 ---	839 (.37)	341 (.15)	0 ---	25 (.01)
4.	$\max_j(p_{ij})$, Discriminant Analysis	210 (.09)	884 (.39)	13 (.01)	17 (.01)	564 (.25)	432 (.19)	9 ---	148 (.07)

correctly classify the same proportion of individuals.

III. Calculating Wage Discrimination

In order to estimate the wage discrimination portion of the general model, we employ a simple linear regression model to explain wage rates. Separate wage equations are estimated for each sex and occupation, and mean values of the independent variables are obtained for the samples used in each estimation (see Table 4). Ordinary least squares is used to estimate the following model for each of the sixteen sex-occupation groups:[12]

$$w_i = a_0 + a_1 TENURE_i + a_2 ED_i + a_3 VOCTRG_i + a_4 IPERS_i + a_5 IESUM_i + a_6 EXPER_i$$

TENURE equals the length of time on one's current job (a variable not often available in data sets used for earnings regressions), and all the other variables are as previously defined. Unfortunately, only a small portion of the variance in wages could be explained in almost every subsample. In addition, several of the subsamples had very few observations because of missing data on the dependent variable. This was particularly a problem for the women's data.

Although our results are far from satisfactory, low R^2's in earnings regressions are the rule rather than the exception. Furthermore, we note that many earnings studies measure experience as current age minus years of schooling, that is as EXPER + TENURE in our notation. We, in contrast, are able to treat general labor force experience and tenure on current job as separate variables that can differentially affect current wages. Thus, we believe the results are adequate to illustrate the procedure developed in section I.

TABLE 4

Mean Values of Variables

OCCUPATION	WAGES (Cents per Hour)		TENURE (Years)		ED (Years)		VOCTRG (Months)	
	M	F	M	F	M	F	M	F
1	491.03	390.68	13.[illegible]	5.29	15.06	14.66	9.56	10.58
2	465.44	349.59	16.16	3.61	12.09	11.96	6.32	8.01
3	316.98	274.21	16.24	4.14	11.62	11.95	6.17	5.16
4	385.82	214.02	11.21	3.15	11.47	10.98	6.22	4.40
5	350.41	263.17	13.58	2.28	9.67	11.28	7.40	2.93
6	273.38	244.05	13.59	4.19	8.98	9.88	3.94	1.84
7	276.77	219.63	8.44	2.19	9.68	10.14	3.85	4.10
8	214.79	188.37	10.21	1.25	7.63	7.37	3.18	1.03

OCCUPATION	IESUM*		IEPERS**		EXPER (Years)	
	M	F	M	F	M	F
1	20.83	22.04	6.79	7.11	15.70	13.84
2	20.68	21.73	7.06	7.14	16.82	19.13
3	22.42	23.19	7.79	8.00	16.18	18.93
4	21.34	23.22	7.09	8.42	22.26	20.85
5	22.08	20.83	7.62	7.28	21.08	20.09
6	23.47	23.94	8.19	8.73	22.48	21.45
7	21.93	23.39	7.68	8.58	27.53	22.47
8	25.52	27.12	9.03	9.62	28.03	24.91

*IESUM can range from 11 to 44. A higher score indicates more positive attitude.

**IEPERS can range from 4 to 16. A higher score indicates stronger motivation.

IV. Allocating the Wage Differential

Thus far we have simulated an occupational distribution of women assuming they face the same occupational attainment structure as men. We have also estimated wage functions for each occupation and sex category. It now remains to decompose the total male-female wage differential using the general model of Section I. However, before specifically identifying the components of the wage differential, consider the following alternative specification of our general model:

$$\bar{w}^M - \bar{w}^F = \sum_j (p_j^M \bar{w}_j^M - p_j^F w_j^F)$$

$$= \Sigma(p_j^M a_j^M - p_j^F a_j^F) + \Sigma(p_j^M x_j^M b_j^M - p_j^F \bar{x}_j^F b_j^F)$$

$$= \underset{(\mathrm{I}^*)}{\Sigma p_j^M (a_j^M - a_j^F)} + \underset{(\mathrm{WAGES}^*)}{\Sigma p_j^M (\bar{x}_j^M b_j^M - \bar{x}_j^F b_j^F)} + \underset{(\mathrm{OCC}^*)}{\Sigma \bar{x}_j^F b_j^F (p_j^M - p_j^F)} \qquad (4^*)$$

$$= \underset{(\mathrm{I}^*)}{\Sigma p_j^M (a_j^M - a_j^F)} + \underset{(\mathrm{WD}^*)}{\Sigma p_j^M \bar{x}_j^M (b_j^M - b_j^F)} + \underset{(\mathrm{PD}^*)}{\Sigma p_j^M (\bar{x}_j^M - \bar{x}_j^F) b_j^F}$$

$$+ \underset{(\mathrm{QD}^*)}{\Sigma \bar{w}_j^F (p_j^M - \hat{p}_j^F)} + \underset{(\mathrm{OD}^*)}{\Sigma \bar{w}_j^F (\hat{p}_j^F - p_j^F)}$$

The notation is exactly the same as in section I, but the interpretation of the components differs. For example, OD* is the portion of the total differential that would be removed if women faced the same occupational attainment structure as men but were paid their current mean occupational wage, whereas OD is the portion that would be removed if the occupational attainment structure were the same but women were paid the men's mean wage. Variables WD and WD* are the portions that would be removed if women faced the same wage determination structure as men, WD, or if men faced the same wage structure as women, WD*.[13]

We now consider the estimates of the components of (4) and (4*) presented in Table 5. From OD*, we see that 23.44 percent of the mean wage differential could be eliminated by ending occupational discrimination alone. Our results indicate, however, that the further elimination of wage discrimination, WD* + I*, would actually increase the earnings gap by about 14 percent. This result may be due to the unreliable coefficient estimates for the wage equations. If, instead of (4*) we look at (4), the proper interpretation of WD + I is that if women faced the same structural relationship between wages and personal characteristics as men in each occupation, the total differential would be reduced by 81.40 - 27.47 = 53.93 percent, or about half. From Table 5 we see that the further removal of occupational discrimination would, according to (4), result in an additional reduction of 47.30 percent, thus eliminating the entire remainder. Productivity differences, PD, and qualification differences, QD, balance each other out.

How do we reconcile these very different alternative estimates? Blinder (197[illegible]), faced with a similar problem for measuring wage discrimination, ignored one of the alternatives, claiming that it was not as easily interpreted as the other. Kahne (1975), on the other hand, published the arithmetic average of the two alternative estimates whenever both were available. The terms are interpretable, but the real problem in trying to allocate portions lies in properly specifying the question to be answered. If the question is, what will happen to the difference between men and women in overall mean wages if women faced the same occupational attainment structure as men?, the answer, assuming no change in the $\bar{w}_j^M$, $\bar{w}_j^F$, is OD*. And if the question is, what will happen to the differential if women face the same wage determination structure as men in every occupation?, the answer, assuming no change in the existing distribution of women, is WD + I. The terms WD* + I* and OD indicate what

TABLE 5

Results for Alternative Specifications of the General Model

	(4) Contribution to Total Wage Differential (Cents Per Hour)	(4) % of Total Differential		(4*) Contribution to Total Wage Differential (Cents Per Hour)	(4*) % of Total Differential
I	-25.90	-27.47	I*	74.03	78.48
WD	76.78	81.40	WD*	-87.41	-92.67
PD	11.75	12.46	PD*	94.74	100.44
QD	-13.13	-13.92	QD*	-9.15	-9.70
OD	44.82	47.30	OD*	22.11	23.44
Total	94.33	100.00	Total	94.33	100.00

would happen to men if they faced the same wage determination or occupational attainment structure as women. Thus the appropriate alternative varies, depending upon the questions asked, and neither specification is necessarily preferable to the other.

Moreover, if we now ask the most important question, If both occupational and wage discrimination were ended simultaneously, what would be the effect on the wage differential?, a different equation provides us with the correct answer:

$$TD = \sum_j \hat{p}_j^F(\bar{x}_j^F b_j^M + a_j^M) - \sum_j p_j^F(x_j^F b_j^F + a_j^F) = 90.18 \tag{6}$$

As indicated in Table 5, the total differential is only 94.33 cents. Thus our model predicts that simultaneously eliminating both wage and occupational discrimination would reduce the overall mean wage differential between men and women to only four cents. This remainder is due to differences between the sexes in wages within occupations resulting from different levels of productivity indicators (education, training, etc.) plus differences in occupational distribution resulting from differing job qualifications. However, unlike equations (4) and (4*), equation (6) cannot be easily decomposed. We need to seek alternatives for allocating the differential between the two types of discrimination. Since the results in (6) are similar to the total differential in (4), we might use the allocation breakdown from (4) which indicates that about half the 'unjustified' differential is due to occupational discrimination. Alternatively, we could allocate the differential by comparing the portions that each component contributes separately to the differential and allocating the portion due to their combined effect in the same proportion. In this case, the percentage due to occupational discrimination would be:

$$\frac{OD^*}{OD^* + (I + WD)} = \frac{23.44}{23.44 + 53.93} = 30\%$$

The remaining 70 percent would be attributable to pure wage discrimination.

These results suggest quite a different breakdown from the standard claims for the importance of occupational discrimination. Gwartney and Stroup (1973), Fuchs (1971), Sanborn (1964), and Sawhill (1973) all suggest that the main difference in mean wages is due to occupational differences. Blinder, on the other hand, attributes only about one-third of the differential to occupational differences, an estimate not far from ours. One possible source of this difference in results is the broad occupational categories used here. To the extent that considerable intraoccupational job segregation occurs, our results understate the effect of occupational segregation. Another factor which could account for the differing conclusions is that these earlier studies measured the extent of occupational segregation by using observed differences in occupational distribution, which certainly overstates the effect of job segregation since, as argued earlier, the mean job qualifications of men and women differ. Certainly, a finer breakdown of occupational categories or a respecification of the wage functions could change the results. Nonetheless, these findings provide the first structural approach to the allocation of wage differentials among their wage and occupation discrimination components.

V. Summary and Conclusions

Many economists have tried to explain existing wage differentials between men and women. This paper has suggested an approach that compares the relative importance of occupational discrimination with that of wage discrimination. Our model allows for variation both in occupational distribution and in wages resulting from differences in job qualifications and productivity indicators. We have demonstrated that the usual approach to wage discrimination is a special case of our general model with some restrictive implicit assumptions. A multinomial logit model has been used to simulate the occupational distribution of women that would exist if they faced the same structure of occupational determination as men. Results indicate that there would be more women in managerial and skilled labor jobs and fewer women clerical and service workers. We then estimated wages as a function of productivity measures for men and women in each occupation so that the components of the wage differential could be calculated. Our results indicate that almost the entire differential could be eliminated by ending both forms of discriminatir with occupational discrimination accounting for one-third to one-half of the differential and pure wage discrimination the remainder.

Appendix Table A

Predicted Occupational Distribution
When Allocated by Highest Probability

Actual Group Membership	Predicted Group Membership 1	2	3	4	5	6	7	8	Total
1	131	48	0	0	27	2	0	0	208
2	66	284	0	0	168	62	0	1	581
3	10	67	0	0	33	7	0	0	117
4	10	46	0	0	48	4	0	0	108
5	13	204	0	0	269	94	0	3	583
6	3	130	0	0	177	102	0	10	422
7	0	32	0	0	59	19	0	3	113
8	0	28	0	0	58	51	0	8	145
Total	233	839	0	0	839	341	0	25	2277

FOOTNOTES

[1]Obviously this list excludes many others. See Kahne (197[illegible]) for a comprehensive review of the literature.

[2]The Schmidt and Strauss model (1975) is the same as (5) but with a different normalization.

[3]The first order conditions are

$$\sum_{i=1}^{N} \frac{e^{x_i'\hat{\gamma}_j}}{\sum_{k=1}^{G} e^{x_i'\hat{\gamma}_k}} x_i = \sum_{i=1}^{N} x_i v_{ij} \qquad j=1, \ldots, G$$

$$\sum_{k=1}^{G} \hat{\gamma}_k = 0$$

[4]Mincer and Polachek (1974) discuss the problems that not having detailed experience information creates in women's earnings functions. Hansen and Weisbrod (1973) argue that, even for men, age is not a good proxy for experience.

[5]The Rotter scale variables are responses of individuals to questions on attitude and motivation. Responses range from one through four, higher responses implying "better" attitude or greater motivation. Alternatively, higher values may be interpreted as representing a greater sense of control over one's position or environment. For a thorough explanation of the Rotter scale, see Andrisani and Nestel (1975).

[6]Since the term in brackets on the left-hand side of the first order conditions (see footnote 3) is just $\hat{p}_{ij}$ we have

$$\sum_i \hat{p}_{ij} \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix} = \sum_i v_{ij} \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix}$$

Hence:

$$\sum_i p_{ij} = \sum_i v_{ij} = N_j \ , \ j = 1, \ldots, J.$$

[7]We recognize that this broad categorization of occupations obscures a great deal of occupational segregation by sex, but we are constrained by computational limitations of the multinomial logit model. For studies of intraoccupational segregation, the reader is referred to Grimm and Stern (1974), Johnson and Stafford (1964), and Malkiel and Malkiel (1973). Part of the distinction, of course, between inter- and intra-occupational segregation depends on how large the occupational categories are.

While the NLS attempts to be a representative national sample, the percentage of men and women in each category deviates slightly from the overall national average.

[8]Although the 34.9 percent figure may be viewed as a measure of the fit of the model, it is an arbitrary one since the correct occupation may frequently have a predicted probability nearly as large. A more serious problem, however, is the lack of individuals assigned to occupations 3, 4, and 7. It is no coincidence that these are the occupations least represented in the sample. The small number of observations in these categories led to the fitting of small coefficients for these groups on the most important variables. Consequently, no individual has a particularly high probability of being in any one of these groups. Table A in the Appendix displays the actual and predicted group memberships from this technique.

[9]Even if we do not wish to make this strong an assumption, the results are of interest since differences between career-oriented men and women in choice of occupation may also be due to discrimination, either because women do not seek jobs they do not feel they will be given a fair chance of getting or because they have been told throughout their lives that women do not belong in certain occupations. Madden (1973) identifies three types of sex discrimination: wage, occupational, and cumulative. These differences in attitude would fall into her third category.

[10]Sample size for the women totalled 2435. Again, the actual occupational breakdowns are similar to, but do not exactly correspond with, national percentages.

[11]In discriminant analysis (DA), classification is done by assigning individuals to the occupation to which they have the highest probability of belonging. Probabilities are computed by assuming that the personal characteristics of individuals are distributed multivariate normal, i.e., for individual i in group g, $x_i \sim N(\mu_g, \Sigma_g)$. This leads to the following expression for probabilities (cf. Eisenbeis and Avery, 1972)

$$p_{ij} = p(y_i = a_j) = \frac{\Pi_j |\Sigma_j|^{-(1/2)} \exp(-\chi^2_{ij}/2)}{\sum_{h=1}^{G} \Pi_h |\Sigma_h|^{-(1/2)} \exp(-\chi^2_{ih}/2)} \qquad (7)$$

where $\chi^2_{i,j} = (x_i - \mu_j)' \Sigma_j^{-1} (x_i - \mu_j)$ and Π_j is the a priori probability of an observation being drawn from group j ($\Pi_j = N_j/N$ for our case). As Eisenbeis and Avery note, the classification rule can also be written as a quadratic discriminant function of the x_i.

[12]The fact that ordinary least squares is used to estimate the wage equations is important, since this insures that $\bar{w}_i = \bar{x}_i b_i + a_i$, a necessary condition for the breakdown as we have written it.

[13]An analogous breakdown can be made for the Blinder model:

$$\bar{x}^H b^H - \bar{x}^L b^L = (\bar{x}^H - \bar{x}^L) b^L + \bar{x}^H (b^H - b^L)$$

Then $(\bar{x}^H - \bar{x}^L) b^L$ is the portion of the earnings differential that would still remain if men faced the same wage structure as women.

BIBLIOGRAPHY

Andrisani, P.J., and G. Nestel (1975). "Internal-External Control and Labor Market Experience," Chapter VI in The Pre-Retirement Years, Vol. 4, U.S. Department of Labor, Manpower Research Monograph No. 15. Washington D.C.: U.S. Government Printing Office.

Bergmann, B. (1971). "The Effect on White Incomes of Discrimination in Employment," Journal of Political Economy. March/April, pp. 294-313.

_______. (1974). "Occupational Segregation, Wages and Profits When Employers Discriminated by Race or Sex," Eastern Economic Journal. April/July, pp. 103-110.

Blau, F.D. (1975). "Sex Segregation of Workers by Enterprise in Clerical Occupations," Chapter 10 in Labor Market Segmentation, R.C. Edwards, et al., eds. Lexington: D.C. Heath & Co.

Blinder, A.S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates," Journal of Human Resources. Fall, pp. 436-455.

Boskin, M.J. (1974). "A Conditional Logit Model of Occupational Choice," Journal of Political Economy. March, pp. 389-898.

Eisenbeis, R.A., and R.B. Avery (1972). Discriminant Analysis and Classification Procedures. Lexington: D.C. Heath & Co.

Fuchs, V.R. (1971). "Differences in Hourly Earnings Between Men and Women," Monthly Labor Review. May, pp. 9-15.

_______. (1974). "Recent Trends and Long-Run Prospects for Female Earnings," American Economic Review. May, pp. 236-242.

Grimm, J.W., and R.N. Stern (1974). "Sex Roles and Internal Labor Market Structures: The 'Female' Semi-Professions," Social Problems. June, pp. 690-1.

Gwartney, J., and R. Stroup (1973). "Measurement of Employment Discrimination According to Sex," Southern Economic Journal. April, pp. 575-87.

Hansen, W.L., and B. Weisbrod (1973). "Human Capital Investment, Schooling, and Earnings: 'The Role of Experience,'" Discussion Paper No. 182-73. Madison: Institute for Research on Poverty.

Johnson, G.E., and F.P. Stafford (1974). "The Earnings and Promotion of Women Faculty," Arerican Economic Review. December, pp. 888-903.

Kahne, H. (1975). "Economic Perspectives on the Roles of Women in the American Economy," Journal of Economic Literature. December, pp. 1249-1292.

Lyle, J.R., and J.L. Ross (1974). Women in Industry: Employment Patterns of Women in Corporate America. Lexington: D.C. Heath & Co.

Madden, J.F. (1973). The Economics of Sex Discrimination. Lexington: D.C. Heath & Co.

Malkiel, B.G., and J.A. Malkiel (1973). "Male-Female Pay Differentials in Professional Employment," American Economic Review. September, pp. 693-705.

Mincer, J., and S. Polachek (1974). "Family Investments in Human Capital: Earnings of Women," Journal of Political Economy. March/April, pp. S76-S108.

Nerlove, M., and S.J. Press (1973). Univariate and Multivariate Log-Linear and Logistic Models, manuscript R-1306-EDA/NIH, Rand Corporation.

Oaxaca, R. (1973). "Sex Discrimination in Wages," in Discrimination in Labor Markets, O. Ashenfelter and A Rees eds. Princeton: Princeton University Press.

Polachek, S.W. (1975). Potential Biases in Measuring Male-Female Discrimination," Journal of Human Resources. Spring, pp. 205-229.

Sanborn, H. (1964). "Pay Differences Between Men and Women," Industrial Labor Relations Review. July, pp. 534-550.

Sawhill, I.V. (1973). "The Economics of Discrimination Against Women: Some New Findings," Journal of Human Resources. Summer, pp. 383-396.

Schmidt, P., and R. Strauss (1975). "The Prediction of Occupation Using Multiple Logit Models," International Economic Review. June, pp. 471-485.

Stevenson, M.H. (1975). "Relative Wages and Sex Segregation by Occupation," Chapter 7 in Sex, Discrimination, and the Division of Labor, C.B. Lloyd ed. New York; Columbia University Press.

Zellner, H. (1972). "Discrimination Against Women, Occupational Segregation, and the Relative Wage," American Economic Review. May, pp. 157-160.

_______. (1975). "The Determinants of Occupational Segregation," Chapter 5 in Sex, Discrimination, and the Division of Labor, C.B. Lloyd ed. New York: Columbia University Press.