THE PREDICTION OF OCCUPATION USING MULTIPLE LOGIT MODELS

By

P. SCHMIDT AND R. P. STRAUSS

Reprinted from

INTERNATIONAL ECONOMIC REVIEW

Vol. 16, No. 2, June, 1975

INTERNATIONAL ECONOMIC REVIEW
Vol. 16, No. 2, June, 1975

THE PREDICTION OF OCCUPATION USING MULTIPLE LOGIT MODELS*

BY PETER SCHMIDT AND ROBERT P. STRAUSS[1]

1. INTRODUCTION

ECONOMIC ANALYSIS OF EMPLOYMENT patterns by race and sex typically has involved construction of indices of relative employment shares for the pertinent race-sex groups in various occupations, given the average level of educational attainment in the occupation. Becker [1] and Gilman [3], among others, have used such normalized percentages to analyze relative access to jobs by blacks and women, as well as to analyze variation in such indices of job discrimination on the basis of other characteristics.

We propose here instead to analyze patterns of employment by estimating a multiple logit model of occupational attainment, using race, sex, educational attainment and labor market experience as explanatory variables. The advantages of this direct approach are that it makes it unnecessary to make assumptions about an appropriate reference point (such as average educational attainment in the occupation), and that the analysis may be based on individual observations rather than percentages.

The plan of the paper is as follows: Section 2 develops the basic model and reports results for 1960, 1967 and 1970. Section 3 extends the model by considering possible regional difference and possible race and sex interactions. Section 4 concludes. The Appendix contains a summary of the multiple logit model and its maximum likelihood estimation.

2. THE PREDICTION OF OCCUPATIONAL LEVEL

Statistical analysis of models with qualitative dependent variables can be viewed as the problem of predicting probabilities for the various possible values (responses) of the dependent variable. Probit and logit analysis are two well-known techniques for the case in which there are only two possible responses—typically the occurrence or non-occurrence of some event.

More recent work has considered the case of an arbitrary number of responses. Most of this work, however, has been concentrated on the case in which the explanatory variables are also qualitative. (See, for example, Goodman [4], Grizzle, Starmer and Koch [5], and the references therein.) Estimation in this

* Manuscript received August 7, 1974; revised November 1, 1974.

[1] We wish to thank the Social Science Data Library, University of North Carolina, for access to the 1960 and 1970 Public Use Samples, The Data and Program Library Service, University of Wisconsin, for an efficient version of the 1967 Survey of Economic Opportunity, and the National Science Foundation and U.S. Department of Labor for their support. We also wish to thank the referees for their helpful comments.

TABLE 1
UNDERLYING TWO DIGIT CENSUS OCCUPATIONS USED IN FIVE-WAY GROUPING

Constructed Title	Component Census Titles
"Professional" (5)	Professional, Technical and Kindred Workers Managers, Officials, and Proprietors except Farm
"White Collar" (4)	Clerical and Kindred Workers Sales Workers
"Craft" (3)	Craftsmen, Foremen, and Kindred Workers Farm Managers and Farmers
"Blue Collar" (2)	Operatives and Kindred Workers Laborers, except Farm and Mine
"Menial" (1)	Private Household workers Service Workers, except Private Household Farm Laborers and Foremen

case is based on the estimated cell frequencies in the underlying multiple-dimension contingency table, and is computationally convenient. However, if certain explanatory variables are inherently continuous, they must be grouped into ranges of values. In some applications this grouping may be a serious disadvantage, since it may introduce an approximation error of unknown magnitude. On the other hand, Theil [**10**] has developed a "multiple logit model" which does allow both an arbitrary number of responses and continuous right hand side variables. The maximum likelihood estimation of this type of model has been discussed by Press [**8**], Nerlove and Press [**7**], and McFadden [**6**].[2]

In this paper we apply the multiple logit model to the prediction of occupation (or type of job) of individuals, based on certain personal characteristics. We begin by examining the effects of four essentially exogenous variables on occupation: race, sex, educational attainment, and labor market experience. Race and sex are zero-one dummies taking on the value one for the more numerous (whites and males) categories. Education is measured in school years, and labor market experience is calculated as age minus years of schooling minus five.

Using these explanatory variables, we predict individuals to be in one of five occupational groups: "professional," "white collar," "craft," "blue collar," and "menial." These five categories are groupings of the major two-digit Census occupational categories, as given in Table 1.

The results obtained here may be of some interest, apart from the prediction of occupation *per se*, because they can also be interpreted as a measure of race and sex discrimination. Non-zero coefficients of the race and sex variables indicate that race and sex affect one's occupation, even when differences in education and experience are taken into account. That is, non-zero coefficients of the race and sex variables indicate differential access to certain occupations, depend-

[2] It may be noted that a linear specification (in which the dependent variable is the probability of being in a particular group) has often been used in empirical work. It is now widely recognized that this approach is deficient for two reasons: first, because of the possibility of predicted probabilities being negative or greater than one; second, because of the difficulties caused by heteroskedasticity of the disturbance. See Nerlove and Press [7] for further discussion.

ing on one's race and sex.

However, it should be emphasized that this interpretation should be made only with some caution. While it is true that race and sex may affect one's occupation because of discrimination by employers, it is also true that race and sex may affect one's preferences for different occupations. In fact, in the absence of discrimination, the present model could be considered to be a model of occupational choice. (See Boskin [**2**] for a similar analysis in a different context.) Ideally, of course, we would like to separate out differences in occupations due to discrimination from differences due to preferences. Unfortunately it is not clear how to do so.

In order to interpret the results presented here as evidence of discrimination, therefore, one has to believe that occupational differences due to preferences are relatively small compared to differences due to discrimination. To these authors, this seems reasonable for differences between races, but less reasonable for differences between sexes.

For the years 1960 and 1970, we analyze 1000 observations drawn from the Public Use Samples. Each sample was drawn randomly from the parent 1 in 1000 sample after the latter had been modified to include only those over 14 years of age, those who were full time workers, and those who had non-zero earnings in the reference year. Excluded, then, are part-time workers, the self-employed, and those who received non-monetary wages. The same procedure was followed for 1967, except that a sample of size 934 was drawn at random from the representative portion of the 1967 Survey of Economic Opportunity.

We estimate, for each year, functions of the form:

$$\log_e (P_2/P_1)_t = \beta_{11} + \beta_{12}\,\text{Education}_t + \beta_{13}\,\text{Experience}_t + \beta_{14}\,\text{Race}_t + \beta_{15}\,\text{Sex}_t$$
$$\log_e (P_3/P_1)_t = \beta_{21} + \beta_{22}\,\text{Education}_t + \beta_{23}\,\text{Experience}_t + \beta_{24}\,\text{Race}_t + \beta_{25}\,\text{Sex}_t$$
$$\log_e (P_4/P_1)_t = \beta_{31} + \beta_{32}\,\text{Education}_t + \beta_{33}\,\text{Experience}_t + \beta_{34}\,\text{Race}_t + \beta_{35}\,\text{Sex}_t$$
$$\log_e (P_5/P_1)_t = \beta_{41} + \beta_{42}\,\text{Education}_t + \beta_{43}\,\text{Experience}_t + \beta_{44}\,\text{Race}_t + \beta_{45}\,\text{Sex}_t$$

where 1 = "menial," 2 = "blue collar," 3 = "craft," 4 = "white collar," and 5 = "professional." We can also derive from these equations the equations for other comparisons. For example, since

$$\log_e (P_4/P_3) = \log_e (P_4/P_1) - \log_e (P_3/P_1)\,,$$

we have:

$$\log_e (P_4/P_3)_t = (\beta_{31} - \beta_{21}) + (\beta_{32} - \beta_{22})\,\text{Education}_t + (\beta_{33} - \beta_{23})\,\text{Experience}_t + (\beta_{34} - \beta_{24})\,\text{Race}_t + (\beta_{35} - \beta_{25})\,\text{Sex}_t\,.$$

The estimated coefficients and their "t ratios" are given in Tables 2–4. The "t ratios" are the ratios of the estimated coefficients to their estimated asymptotic standard errors, and are asymptotically distributed as $N(0, 1)$ under the null hypothesis that the associated coefficients are zero.

TABLE 2

COEFFICIENTS AND "t RATIOS", 1960[1]

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	1.293	.1238	.02432	1.244	.7988
	(2.18)	(−2.71)	(−2.74)	(4.46)	(3.23)
$\log_e (P_3/P_1)$	−4.086	.0490	−.00957	2.747	2.138
	(−4.56)	(0.92)	(−0.94)	(5.02)	(5.34)
$\log_e (P_4/P_1)$	−3.358	.2163	−.01682	2.8517	−.8087
	(−4.05)	(4.17)	(−1.70)	(5.11)	(−3.14)
$\log_e (P_5/P_1)$	−6.025	.4128	−.00126	1.879	.2263
	(−7.11)	(7.59)	(−0.12)	(3.83)	(0.80)
$\log_e (P_3/P_2)$	−5.379	.1728	.01475	1.5030	1.3395
	(−6.78)	(4.11)	(1.85)	(2.85)	(3.61)
$\log_e (P_4/P_2)$	−4.651	.3401	.00750	1.6074	−1.6075
	(−6.26)	(7.94)	(0.95)	(2.93)	(−7.62)
$\log_e (P_5/P_2)$	−7.318	.5366	.02306	.6347	−.5725
	(−9.09)	(11.66)	(2.72)	(1.32)	(−2.33)
$\log_e (P_4/P_3)$	.7280	.1673	−.00725	.1044	−2.947
	(0.76)	(3.58)	(−0.80)	(0.14)	(−8.00)
$\log_e (P_5/P_3)$	−1.939	.3638	.00831	−.8638	−1.912
	(−2.05)	(7.58)	(0.89)	(−1.31)	(−4.93)
$\log_e (P_5/P_4)$	−2.667	.1965	.01556	−.9727	1.035
	(−3.11)	(4.69)	(1.86)	(−1.45)	(4.66)

[1] Group 5 = "professional," 4 = "white collar," 3 = "craft," 2 = "blue collar," 1 = "menial."

TABLE 3

COEFFICIENTS AND "t RATIOS," 1970

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	1.056	−.1239	−.01491	.7000	1.2519
	(1.56)	(−2.53)	(−1.75)	(2.32)	(5.15)
$\log_e (P_3/P_1)$	−3.769	−.0014	.00776	1.4575	3.1117
	(−4.18)	(−0.03)	(0.80)	(3.35)	(6.56)
$\log_e (P_4/P_1)$	−3.305	.2254	.00300	1.7624	−.5233
	(−4.22)	(4.27)	(0.35)	(4.27)	(−2.22)
$\log_e (P_5/P_1)$	−5.959	.4292	.00779	.9758	.6557
	(−7.51)	(7.96)	(0.88)	(2.62)	(2.65)
$\log_e (P_3/P_2)$	−4.825	.1225	.02268	.7575	1.8598
	(−6.14)	(2.86)	(2.85)	(1.90)	(4.02)
$\log_e (P_4/P_2)$	−4.361	.3493	.01792	1.0624	−1.775
	(−6.25)	(7.69)	(2.43)	(2.60)	(−8.35)
$\log_e (P_5/P_2)$	−7.015	.5531	.02270	.2758	−.5962
	(−10.03)	(11.89)	(3.02)	(0.77)	(−2.62)
$\log_e (P_4/P_3)$	.4640	.2268	−.00476	.3049	−3.635
	(0.52)	(4.57)	(−0.56)	(0.56)	(−7.97)
$\log_e (P_5/P_3)$	−2.190	.4306	.00002	−.4817	−2.456
	(−2.49)	(8.71)	(0.00)	(−1.03)	(−5.30)
$\log_e (P_5/P_4)$	−2.654	.2038	.00478	−.7866	1.179
	(−3.77)	(4.95)	(0.67)	(−1.77)	(5.96)

TABLE 4

COEFFICIENTS AND "*t* RATIOS," 1967

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	1.203 (1.71)	.0806 (−1.68)	−.0263 (−2.54)	.6695 (2.02)	1.1793 (4.62)
$\log_e (P_3/P_1)$	−4.319 (−4.45)	.1153 (2.13)	−.00662 (−0.57)	.8181 (1.90)	3.793 (6.69)
$\log_e (P_4/P_1)$	−2.821 (−3.37)	.2741 (5.05)	−.03341 (−3.09)	1.778 (4.11)	−.5970 (−2.23)
$\log_e (P_5/P_1)$	−6.098 (−7.04)	.4773 (8.56)	−.01242 (−1.12)	1.066 (2.59)	1.077 (3.85)
$\log_e (P_3/P_2)$	−5.522 (−6.60)	.1959 (4.64)	.01969 (2.26)	.1468 (0.39)	2.6137 (6.60)
$\log_e (P_4/P_2)$	−4.024 (−5.57)	.3547 (7.72)	−.00710 (−0.83)	1.1053 (2.68)	−1.7763 (−8.03)
$\log_e (P_5/P_2)$	−7.301 (−9.87)	.4479 (11.89)	.01389 (1.62)	.3965 (1.04)	−.1023 (−0.43)
$\log_e (P_4/P_3)$	1.498 (1.56)	.1588 (3.21)	−.02679 (−2.72)	.9567 (1.94)	−4.390 (−8.00)
$\log_e (P_5/P_3)$	−1.779 (−1.92)	.3620 (7.63)	−.00580 (−0.62)	.2479 (0.56)	−2.716 (−4.90)
$\log_e (P_5/P_4)$	−3.277 (−4.34)	.2032 (4.68)	.02099 (2.46)	−.7088 (−1.63)	1.674 (7.66)

Let us consider first the effects of education. The preponderance of positive coefficients, for all three years, indicates that more education makes it more likely to be in a higher-numbered group as opposed to a lower numbered group. Presumably this is what one expects—education enables one to move "up" the job scale. The only exception is that, in all three years, more education makes it *less* likely to be in a "blue collar" position than in a "menial" position. Since we are using only formal education as our measure of extent of training, we may be obtaining spurious inferences about the effect of training on occupational attainment. This might be the case for the "blue collar" versus "menial" comparison, for the "blue collar" occupations tend to contain a fairly high component of non-formal training (e.g., on the job training).

The effects of experience are somewhat puzzling. The four coefficients that are significant in all three years indicate that more experience makes it more likely to be in a "menial" occupation versus a "blue collar" occupation, more likely to be in a "craft" occupation versus a "blue collar" occupation, more likely to be in a "professional" occupation versus a "blue collar" occupation, and more likely to be in a "professional" versus a "white collar" occupation. Put another way, this says that blue collar workers tend to be young and professionals tend to be old, holding race, sex, and educational attainment constant.

The effects of sex are fairly clear-cut. The results for all three are basically the same and say the following. If we order the occupations as follows:

"white collar"
"menial"

"professional"
"blue collar"
"craft",

then, other things held constant, being female makes one more likely to be in any occupational group relative to any other occupational group *lower* on the list. Conversely, being male makes it more likely to be in any group relative to any other group *higher* on the list. So, for example, being female makes it more likely to be in the "white collar" occupation, relative to any other occupation. Similarly, being female makes it more likely to be in the "menial" group, relative to any other group except "white collar," and so forth.

Note that the "professional" group is in the middle; sex has relatively little effect on entry into this group. Females are apt to end up in "white collar" or "menial" positions, while males are apt to end up in "blue collar" or "craft" positions.

Finally, let us consider the effect of race on occupation. If we order the occupations as follows:

"menial"
"blue collar"
"craft," "professional"
"white collar",

then being *black* makes it more likely to be in any group relative to any other group *lower* on the list. Conversely, being *white* makes it more likely to be in any group relative to any other group *higher* on the list. The fact that the "craft" and "professional" groups are on the same line indicates that the effect of race on the odds of being in the "craft" group relative to the "professionl" group was statistically insignificant, and varied in sign over the three years.

Essentially, what these results show is that being black makes it more likely to be in one of the less desirable groups—"menial" and "blue collar." Furthermore, the worst discrimination is encountered in "white collar" positions, not, as some might have expected, in the "craft" positions.

In order to see in another way what these results are saying, we have also evaluated the probabilities of being in each of the five occupations for each of the three years. These probabilities were evaluated at the sample means for education and experience, and for all four permutations of race and sex, using formula (2) of the Appendix. The results are given in Table 5, and bear out much of what was said above. Note, for example, the preponderance of black females in the "menial" group, of white females in the "white collar" group, and of males (especially white males) in the "craft" group.

A final thing worth noting is the intertemporal change, 1960–1970, in the coefficients of the race and sex variables. A movement of these coefficients toward zero represents a decrease in discrimination, and conversely. With this in mind, it is interesting to note that the coefficient of the race variable decreased from 1960 to 1970 in nine out of ten equations, while the coefficient of the sex variable increased in nine of ten equations. This, then, would indicate a decrease in occupational differences due to race, but an increase in occupational

TABLE 5

PROBABILITIES OF BEING IN EACH OCCUPATION, GIVEN AVERAGE EDUCATION AND EXPERIENCE

	Race-Sex Combination	Menial	Blue Collar	Craft	White Collar	Professional
1960[1]	Black female	.479	.280	.012	.124	.104
	Black male	.345	.448	.073	.040	.094
	White female	.107	.217	.041	.481	.153
	White male	.080	.359	.261	.159	.142
1970[2]	Black female	.396	.188	.011	.219	.187
	Black male	.222	.368	.136	.073	.202
	White female	.153	.146	.018	.492	.192
	White male	.089	.296	.232	.169	.214
1967[3]	Black female	.366	.258	.015	.221	.140
	Black male	.151	.346	.283	.050	.170
	White female	.140	.192	.013	.499	.156
	White male	.067	.299	.284	.131	.219

[1] Average education = 10.742; average experience = 24.570.
[2] Average education = 11.718; average experience = 23.406.
[3] Average education = 11.400; average experience = 24.147.

differences due to sex. The reader is warned, however, that none of the inter-temporal (1960–1970) changes in these coefficients is significant at the 5% level.

3. SOME EXTENSIONS OF THE MODEL

The model used above is an admittedly simple one. In this section we test its adequacy by considering certain extensions (or complications).

The first problem we deal with is the fact that occupational patterns may vary substantially over various regions, beyond that explainable in terms of regional differences in our four explanatory variables. To see if taking this into account would substantially change our conclusions, we obtained random samples of size 1000 for each of the four Census regions (Northeast, North Central, South, West) from the representative portion of the 1967 Survey of Economic Opportunity. The model of the previous section was then estimated for each region. These results are given in Tables 6–9, and the corresponding predicted probabilities (for average age and experience) are given in Table 10.

Coefficients which are significantly different (at the 5% level) from those of the national sample (as given in Table 4) are marked with an asterisk. As is clear from glancing at Tables 6–9, there are relatively few such significant changes. A more detailed analysis of the regional results will be left to the reader, since our main interest in them was just to verify that the inter-regional differences were not so great as to cast doubt on the usefulness of the national sample.

As the reader may easily observe, the predicted probabilities in Table 10 do

TABLE 6

COEFFICIENTS AND "*t* RATIOS," 1967, NORTH EAST REGION

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	1.649	−.0887	−.01915	.4549	.6291
	(2.27)	(−1.83)	(−2.05)	(1.31)	(2.57)
$\log_e (P_3/P_1)$	−3.069	.0492	−.00992	.8653	2.998
	(−3.26)	(0.90)	(−0.93)	(1.92)	(6.23)
$\log_e (P_4/P_1)$	−1.908	.2504	−.01611	.7835	−.6973
	(−2.44)	(4.78)	(−1.67)	(2.02)	(−2.76)
$\log_e (P_5/P_1)$	−5.252	.4463	−.00979	1.023	.3897
	(−6.37)	(8.36)	(−0.99)	(2.40)	(1.49)
$\log_e (P_3/P_2)$	−4.716	.1379	.00922	.4104	2.368
	(−5.90)	(3.23)	(1.10)	(1.04)	(5.20)
$\log_e (P_4/P_2)$	−3.557	.3391	.00303	.3286	−1.326
	(−5.63)	(8.03)	(0.41)	(0.97)	(−6.58)
$\log_e (P_5/P_2)$	−6.901	.5350	.00936	.5681	−.2394
	(−10.14)	(12.32)	(1.22)	(1.50)	(−1.13)
$\log_e (P_4/P_3)$	1.159	.2012	−.00619	−.0818	−3.695
	(1.36)	(4.29)	(0.68)	(−0.18)	(−8.07)
$\log_e (P_5/P_3)$	−2.185	.3971	.00014	.1577	−2.608
	(−2.53)	(8.62)	(0.02)	(0.34)	(−5.65)
$\log_e (P_5/P_4)$	−3.334	.1959	.00633	.2395	1.087*
	(−5.23)	(5.19)	(0.86)	(0.60)	(5.54)

TABLE 7

COEFFICIENTS AND "*t* RATIOS" 1967, NORTH CENTRAL REGION

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	.3810	.01150	−.01803	.4720	1.2200
	(0.48)	(0.21)	(−1.78)	(1.32)	(4.82)
$\log_e (P_3/P_1)$	−4.822	.0971	.00249	1.2970	3.742
	(−4.40)	(1.53)	(0.21)	(2.41)	(6.54)
$\log_e (P_4/P_1)$	−3.116	.3145	−.01423	1.0169	.0730
	(−3.63)	(5.33)	(−1.37)	(2.40)	(0.28)
$\log_e (P_5/P_1)$	−6.440	.5192	.00179	.7437	1.160
	(−7.07)	(8.50)	(0.17)	(1.66)	(4.21)
$\log_e (P_3/P_2)$	−5.203	.0856	.02052	.8250	2.5219
	(−5.65)	(1.83)	(2.34)	(1.73)	(4.60)
$\log_e (P_4/P_2)$	−3.497	.3030	.00380	.5449	−1.147*
	(−5.29)	(7.01)	(0.50)	(1.46)	(−5.87)
$\log_e (P_5/P_2)$	−6.821	.5077	.01982	.2717	−.0160
	(−9.65)	(11.32)	(2.52)	(0.70)	(−0.28)
$\log_e (P_4/P_3)$	1.706	.2174	−.01672	−.2801	−3.669
	(1.72)	(4.25)	(−1.76)	(−0.51)	(−6.70)
$\log_e (P_5/P_3)$	−1.618	.4221	−.00070	−.5533	−2.582
	(−1.62)	(8.24)	(−0.07)	(−1.00)	(−4.65)
$\log_e (P_5/P_4)$	−3.324	.2047	.01602	−.2732	1.087*
	(−4.78)	(5.08)	(2.06)	(−0.63)	(5.44)

TABLE 8

COEFFICIENTS AND "t RATIOS," SOUTHERN REGION

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	.1140	−.0250	−.01950	1.073	1.379
	(0.20)	(−0.61)	(−2.09)	(4.41)	(5.93)
$\log_e (P_3/P_1)$	−5.324	.0832	−.00251	2.373*	4.105
	(−6.47)	(1.74)	(−0.23)	(6.77)	(8.51)
$\log_e (P_4/P_1)$	−4.538	.2936	−.01204	2.5615	.1682*
	(−6.12)	(6.07)	(−1.19)	(6.54)	(0.64)
$\log_e (P_5/P_1)$	−7.038	.4855	.00948	1.809	1.125
	(−9.44)	(9.90)	(0.93)	(5.50)	(4.25)
$\log_e (P_3/P_2)$	−5.438	.1082	.01699	1.3001*	2.726
	(−7.64)	(2.90)	(1.99)	(4.22)	(5.97)
$\log_e (P_4/P_2)$	−4.652	.3186	.00746	1.4885	−1.2113
	(−7.19)	(7.75)	(0.89)	(3.95)	(−5.71)
$\log_e (P_5/P_2)$	−7.152	.5105	.02898	.7360	−.2545
	(−11.18)	(12.28)	(3.48)	(2.41)	(−1.19)
$\log_e (P_4/P_3)$	.7860	.2104	−.00953	.1884	−3.937
	(0.94)	(4.74)	(−0.99)	(0.42)	(−8.50)
$\log_e (P_5/P_3)$	−1.714	.4023	.01199	−.5641	−2.980
	(−2.12)	(9.26)	(1.30)	(−1.46)	(−6.43)
$\log_e (P_5/P_4)$	−2.500	.1919	.02152	−.7525	.9568*
	(−3.71)	(5.03)	(2.60)	(−1.81)	(4.58)

TABLE 9

COEFFICIENT AND "t RATIOS," 1967 WESTERN REGION

Dependent Variable	Constant	Education	Experience	Race	Sex
$\log_e (P_2/P_1)$	−.0450	.00610	−.02170	.1961	1.3691
	(−0.06)	(0.13)	(−2.11)	(0.44)	(4.76)
$\log_e (P_3/P_1)$	−3.944	.06220	−.01882	1.6269	3.2798
	(−3.62)	(1.23)	(−1.76)	(2.36)	(6.31)
$\log_e (P_4/P_1)$	−2.940	.2309	−.02107	2.1554	−.7702
	(−3.23)	(4.59)	(−2.13)	(3.67)	(−3.04)
$\log_e (P_5/P_1)$	−6.116	.4763	−.00439	1.184	.5908
	(−6.98)	(9.32)	(−0.44)	(2.45)	(2.27)
$\log_e (P_3/P_2)$	−3.899	.05610*	.00288	1.4308	1.9107
	(−3.93)	(1.35)	(0.32)	(2.23)	(3.69)
$\log_e (P_4/P_2)$	−2.895	.2248*	.00063	1.9593	−2.1393
	(−3.43)	(5.07)	(0.07)	(3.40)	(−8.62)
$\log_e (P_5/P_2)$	−6.071	.4702	−.01731	.9879	−.7783
	(−7.85)	(10.76)	(−2.08)	(2.24)	(−3.06)
$\log_e (P_4/P_3)$	1.004	.1687	−.00225	.5285	−4.0500
	(0.91)	(3.72)	(−0.25)	(0.67)	(−8.17)
$\log_e (P_5/P_3)$	−2.172	.4141	.01443	−.4429	−2.689
	(−2.14)	(9.53)	(1.69)	(−0.66)	(−5.39)
$\log_e (P_5/P_4)$	−3.176	.2454	.01668	−.9714	1.361
	(−3.95)	(6.38)	(2.27)	(−1.67)	(7.15)

TABLE 10

PROBABILITIES OF BEING IN EACH OCCUPATION, GIVEN AVERAGE EDUCATION AND EXPERIENCE

Race-Sex Combination	Region	"Menial"	"Blue Collar"	"Craft"	"White Collar"	"Professional"
Black female	US[1]	.366	.258	.015	.221	.140
	NE[2]	.211	.245	.014	.379	.150
	NC[3]	.257	.277	.007	.296	.164
	S[4]	.515	.273	.006	.098	.108
	W[5]	.349	.217	.009	.190	.234
Black male	US	.151	.346	.283	.050	.170
	NE	.156	.340	.201	.139	.164
	NC	.111	.405	.122	.137	.226
	S	.215	.452	.147	.048	.139
	W	.180	.435	.126	.045	.216
White female	US	.140	.192	.013	.499	.156
	NE	.112	.206	.017	.442	.223
	NC	.136	.235	.013	.433	.183
	S	.156	.241	.019	.384	.200
	W	.114	.086	.015	.535	.250
White male	US	.067	.299	.284	.131	.219
	NE	.081	.278	.247	.158	.236
	NC	.054	.315	.216	.184	.231
	S	.047	.288	.343	.136	.185
	W	.073	.217	.263	.158	.289

[1] Average education = 11.509, average experience = 24.203
[2] Average education = 11.552, average experience = 24.730
[3] Average education = 11.458, average experience = 24.245
[4] Average education = 10.807, average experience = 24.571
[5] Average education = 12.219, average experience = 23.265.

show obvious variation across regions. This is due both to the interregional differences in estimated coefficients and to differences in average education and experience.

The second problem with which we deal is that our specification (using race and sex dummies) allows only the constant term to be shifted by race and sex. That is, the coefficients of the education and experience variables are implicitly assumed to be the same for all people, regardless of race and sex. Also the coefficient of the race dummy is implicitly assumed to be the same for both sexes, and vice-versa.

This implicit assumption may be untrue for a number of reasons. For example, one could argue that the coefficient of the experience variable should not be the same for females as for males. As the reader may recall, experience was defined

TABLE 11

COEFFICIENTS AND "t RATIOS," 1967, SAMPLE OF MALES ONLY

Dependent Variable	Constant	Education	Experience	Race
$\log_e (P_2/P_1)$	3.524 (4.27)	−.03437 (−0.63)	−.02746 (−2.23)	−1.066 (−2.16)
$\log_e (P_3/P_1)$	.01792 (0.02)	.08159 (1.47)	.00266 (0.21)	.4610 (0.82)
$\log_e (P_4/P_1)$	−2.489 (−2.55)	.3003 (4.99)	−.00740 (−0.56)	.2981 (0.48)
$\log_e (P_5/P_1)$	−3.694 (−3.91)	.4487 (7.59)	−.01234 (−0.96)	.1675 (0.29)
$\log_e (P_3/P_2)$	−3.506 (−6.27)	.1160 (3.23)	.03012 (3.82)	1.527 (4.45)
$\log_e (P_4/P_2)$	−6.013 (−8.56)	.3347 (7.81)	.02007 (2.21)	1.364 (3.21)
$\log_e (P_5/P_2)$	−7.218 (−10.95)	.4831 (11.67)	.05120 (17.76)	1.233 (3.34)
$\log_e (P_4/P_3)$	−2.507 (−3.39)	.2187 (5.24)	−.01006 (−1.11)	−.1630 (−0.32)
$\log_e (P_5/P_3)$	−3.712 (−5.36)	.3671 (9.24)	−.01500 (−1.77)	−.2936 (−0.64)
$\log_e (P_5/P_4)$	−1.205 (−1.67)	.1484 (3.90)	−.00494 (−0.55)	−.1306 (−0.26)

as age minus education minus five, and should accurately measure actual labor market experience for individuals who began school at age five and worked continuously since leaving school. This raises the possibility that we have systematically overstated the experience of females, since females (specially married females) are more apt to periodically leave the labor force than males are. Furthermore, even if experience were accurately measured, it is not entirely clear that its effects on the occupational structure of females would be the same as its effects on the occupational structure of males.

As another example of possible interactions of explanatory variables, it could be argued that the coefficient of the education variable should not be the same for blacks as for whites. If blacks receive (or have received in the past) an inferior education compared to whites who have attended school the same number of years, and if employers take this into account, then we have systematically overstated the "actual" education of blacks.

In order to check on these possibilities, we have repeated the above analysis on samples stratified by sex and race. That is, we first generated random samples of 1000 females and 1000 males from the representative portion of the 1967 Survey of Economic Opportunity. On each sample we ran the model of the previous section, except, of course, without the sex dummy. These results are given in Tables 11 and 12. We then generated random samples of 1000 blacks and 1000 whites and ran the previous model, except without the race dummy. These results are given in Tables 13 and 14.

Consider first the results of the stratification by sex (Tables 11–12). As the

TABLE 12

COEFFICIENTS AND "t RATIOS," 1967, SAMPLE OF FEMALES ONLY

Dependent Variable	Constant	Education	Experience	Race
$\log_e (P_2/P_1)$	.2285 (0.36)	−.04296 (−0.92)	−.02792 (−2.96)	1.630 (5.98)
$\log_e (P_3/P_1)$	−3.845 (−2.07)	.01518 (0.12)	−.01814 (−0.75)	2.098 (1.96)
$\log_e (P_4/P_1)$	−3.389 (−5.01)	.2871 (5.89)	−.04040 (−4.51)	2.538 (8.24)
$\log_e (P_5/P_1)$	−6.755 (−8.55)	.5359 (9.54)	−.02622 (−2.66)	1.810 (5.39)
$\log_e (P_3/P_2)$	−4.073 (−2.21)	.05814 (0.46)	.00978 (0.41)	.4675 (0.43)
$\log_e (P_4/P_2)$	−3.617 (−5.85)	.3300 (7.56)	−.01247 (−1.71)	.9077 (2.78)
$\log_e (P_5/P_2)$	−6.983 (−9.41)	.5788 (11.06)	.00170 (0.20)	.1795 (0.51)
$\log_e (P_4/P_3)$	.4564 (0.25)	.2719 (2.17)	−.02225 (−0.95)	.4402 (0.41)
$\log_e (P_5/P_3)$	−2.910 (−1.54)	.5207 (4.05)	−.00808 (−0.34)	−.2880 (−0.26)
$\log_e (P_5/P_4)$	−3.366 (−5.38)	.2488 (6.21)	.01418 (2.07)	−.7282 (−2.12)

TABLE 13

COEFFICIENTS AND "t RATIOS," 1967, SAMPLE OF WHITES ONLY

Dependent Variable	Constant	Education	Experience	Sex
$\log_e (P_2/P_1)$	2.013 (2.62)	−.05893 (−1.08)	−.03266 (−2.97)	.7499 (2.75)
$\log_e (P_3/P_1)$	−2.965 (−2.50)	.01799 (0.31)	−.01620 (−1.37)	4.459 (4.98)
$\log_e (P_4/P_1)$	−1.537 (−1.92)	.3103 (5.44)	−.01570 (−1.42)	−.8655 (−3.20)
$\log_e (P_5/P_1)$	−5.231 (−6.18)	.5318 (8.99)	−.00737 (−0.65)	.4225 (1.49)
$\log_e (P_3/P_2)$	−4.979 (−4.75)	.07692 (1.78)	.01646 (1.88)	3.709 (4.21)
$\log_e (P_4/P_2)$	−3.550 (−5.81)	.3692 (8.19)	.01695 (2.06)	−1.615 (−7.65)
$\log_e (P_5/P_2)$	−7.244 (−10.92)	.5907 (12.52)	.02528 (3.00)	−.3275 (−1.44)
$\log_e (P_4/P_3)$	1.429 (1.31)	.2923 (5.96)	.00050 (0.05)	−5.324 (−6.06)
$\log_e (P_5/P_3)$	−2.266 (−2.06)	.5138 (10.38)	.00883 (0.95)	−4.036 (−4.57)
$\log_e (P_5/P_4)$	−3.694 (−6.31)	.2215 (5.68)	.00833 (1.07)	1.288 (6.63)

TABLE 14

COEFFICIENTS AND "t RATIOS," 1967, SAMPLE OF BLACKS ONLY

Dependent Variable	Constant	Education	Experience	Sex
$\log_e (P_2/P_1)$	−.1027 (−0.22)	.00966 (0.30)	−.03320 (−4.16)	2.249 (12.15)
$\log_e (P_3/P_1)$	−4.876 (−5.65)	.1217 (2.56)	−.01097 (−0.91)	4.087 (7.65)
$\log_e (P_4/P_1)$	−3.885 (−5.99)	.3188 (6.99)	−.03088 (−3.04)	1.151 (4.82)
$\log_e (P_5/P_1)$	−6.804 (−8.24)	.4971 (8.67)	−.02015 (−1.58)	.8764 (2.90)
$\log_e (P_3/P_2)$	−4.773 (−5.91)	.1120 (2.64)	.02223 (2.04)	1.838 (3.47)
$\log_e (P_4/P_2)$	−3.783 (−6.14)	.3091 (7.14)	.00232 (0.24)	−1.098 (−4.98)
$\log_e (P_5/P_2)$	−6.702 (−8.34)	.4875 (8.74)	.01305 (1.04)	−1.373 (−4.80)
$\log_e (P_4/P_3)$	.9904 (1.07)	.1971 (3.65)	−.01990 (−1.52)	−2.936 (−5.38)
$\log_e (P_5/P_3)$	−1.929 (−1.82)	.3755 (5.84)	−.00918 (−0.60)	−3.211 (−5.59)
$\log_e (P_5/P_4)$	−2.919 (−3.53)	.1783 (3.18)	.01072 (0.80)	−.2749 (−0.91)

reader can verify, the differences in the coefficients of education, experienec and the race dummy are relatively small. Of particular interest is the fact that the coefficients of the experience variable are *not* systematically smaller for females than for males, as might have been expected if we had (as explained above) actually overstated the experience of females relative to males. While this does not necessarily imply that we have correctly measured experince, it does suggest that male-female differences, for whatever reason, are not so large as to destroy the usefulness of the joint sample.

Similar comments apply to the results of the stratification by race (Tables 13–14). The differences in the coefficients of education, experience and the sex dummy are not terribly large. In this case it is true in 7 cases out of 10 that the coefficient of the education variable is smaller for blacks than for whites, as might be expected if we had (as explained above) overstated the "actual" education of blacks. However, only two of the changes in coefficients were significant at the 5% level. Again, we conclude that the differences are not sufficiently large to draw into question the usefulness of the joint sample.

From the results on the samples stratified by race and sex, it is again possible to generate the predicted probabilities that an individual be in each occupation, given average education and experience. These probabilities are given in Table 15. For ease of comparison the original 1967 probabilities, given in Table 5, are recopied here as well. The probabilities based on the race of sex stratified samples are roughly similar to those based on the original sample; they differ more for blacks than for whites.

TABLE 15

PROBABILITIES OF BEING IN EACH OCCUPATION, GIVEN AVERAGE EDUCATION AND EXPERIENCE[1]

Race-Sex Combination	"Menial"	"Blue Collar"	"Craft"	"White Collar"	"Professional"
Black female[2]	.366	.258	.015	.221	.140
Black male[2]	.151	.346	.283	.050	.170
White female[2]	.140	.192	.013	.499	.156
White male[2]	.067	.299	.284	.131	.219
Black female[3]	.494	.194	.008	.166	.138
Black male[4]	.048	.569	.132	.102	.148
White female[3]	.110	.220	.015	.468	.187
White male[4]	.063	.255	.274	.180	.228
Black female[5]	.490	.221	.011	.181	.097
Black male[5]	.120	.514	.168	.140	.057
White female[6]	.102	.178	.004	.518	.197
White male[6]	.074	.274	.275	.158	.218

[1] Average education and experience as given in Table 5.
[2] Based on sample of 934 individuals of both races and sexes.
[3] Based on sample of 1000 females only.
[4] Based on sample of 1000 males only.
[5] Based on sample of 1000 blacks only.
[6] Based on sample of 1000 whites only.

4. CONCLUSION

This paper has analyzed the occupation of people in terms of their education, experience, race and sex. Race and sex were found to have strong effects. That is, among people of equal education and experience, race and sex strongly influence what sort of a job these people obtain. This can be interpreted as evidence of race and sex discrimination, since it argues that existing patterns of employment cannot be explained merely by black-white or male-female differences in education and experience.

University of North Carolina, U.S.A.

APPENDIX

Let there be N possible responses, with probabilities $P_1, P_2, \ldots, P_N$. Then the multiple logit model can be written as:

$$(1) \qquad \log_e\left[\frac{P_{jt}}{P_{1t}}\right] = X_t\beta_j, \qquad j = 2, 3, \ldots, N; \; t = 1, 2, 3, \ldots, T;$$

where t is the observation index, T the number of observations, X_t the t'-th observation on a $1 \times K$ vector of explanatory variables, and β_j a $K \times 1$ vector of (unknown) parameters.

The $N - 1$ equations in (1), plus the requirement that the probabilities for every t sum to one, determine the probabilities uniquely. Explicitly, the solution is:

$$P_{1t} = \frac{1}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}}$$

(2)

$$P_{it} = \frac{e^{X_t \beta_i}}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}} \qquad i = 2, \ldots, N.$$

This model can be estimated by maximum likelihood by observing that the likelihood function is

(3)
$$L = \prod_{t \in \theta_1} P_{t1} \prod_{t \in \theta_2} P_{t2} \cdots \prod_{t \in \theta_N} P_{tN}$$

where

(4)
$$\theta_j = \{t \mid j\text{-th response is observed}\} .$$

Hence

(5)
$$\begin{aligned} L &= \prod_{t \in \theta_1} \frac{1}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}} \prod_{i=2}^{N} \prod_{t \in \theta_i} \frac{e^{X_t \beta_i}}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}} \\ &= \prod_{t=1}^{T} \left\{ \frac{1}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}} \right\} \prod_{i=2}^{N} \prod_{t \in \theta_i} e^{X_t \beta_i} . \end{aligned}$$

The maximization of this, or its logarithm, can be done using a non-linear maximization program.

To get asymptotic variances for the estimates of $\beta_2, \beta_3, \ldots, \beta_N$, it is necessary to form the information matrix. This turns out to be of the form

(6)
$$\mathscr{I} = \begin{bmatrix} \mathscr{I}_{22} & \mathscr{I}_{23} & \cdots & \mathscr{I}_{2N} \\ \mathscr{I}_{32} & \mathscr{I}_{33} & \cdots & \mathscr{I}_{3N} \\ \vdots & & & \vdots \\ \mathscr{I}_{N2} & \mathscr{I}_{N3} & \cdots & \mathscr{I}_{NN} \end{bmatrix}$$

where

(7)
$$\mathscr{I}_{rr} = \sum_{t=1}^{T} P_{rt}(1 - P_{rt}) X'_t X_t \qquad r = 2, \ldots, N$$

$$\mathscr{I}_{rs} = - \sum_{t=1}^{T} P_{rt} P_{st} X'_t X_t \qquad r, s = 2, \ldots, N, \; r \neq s.$$

The inverse of $\mathscr{I}$ is then the asymptotic covariance matrix of

$$\hat{\beta} = (\hat{\beta}'_2, \hat{\beta}'_3, \ldots, \hat{\beta}'_N)' .$$