1985

"New and Improved" Estimates of Qualification Discrimination

Joni Hersch

Joe A. Stone

Communications

"New and Improved" Estimates of Qualification Discrimination

I. Introduction

What is meant by employment discrimination? The traditional answer to this question is that discrimination is the difference in wages related to sex or race for people with equal productivity, where in practice productivity is measured by individual job qualifications and personal attributes associated with productivity. In the case of sex discrimination, wage regressions of this sort invariably indicate that women earn less than men with equivalent measured qualifications. Recently, however, an alternative view of employment discrimination has been offered by Kamalich and Polachek [8], Conway and Roberts [2], and others. In this view discrimination is the difference in productivity related to sex or race for people with equal wages, where in practice productivity is again measured by individual job qualifications and personal attributes. Unlike the empirical results for wage discrimination, results to date for regressions of qualifications on wages (sometimes referred to as "reverse" regressions) indicate no significant discrimination in qualifications for men and women earning the same wage.

The explanation for differences between the two empirical measures is related to a number of issues, e.g., whether individual qualifications and attributes are simply indicators or true determinants of productivity, whether there are omitted variables or measurement errors correlated with sex, and whether such omissions and errors are more serious for wages or qualifications [1; 2; 3; 4; 5; 8; 12]. In fact in the absence of any sources of error in the relationships among wages and productivity, wage and qualification discrimination are necessarily the same. Although the relative merits of wage versus qualification regressions are a subject of great controversy, the objective of this paper is simply to suggest an improved method of performing qualification regressions and to present new estimates of qualification discrimination for men and women. The fundamental deficiency of previous methods is their failure to treat multiple qualifications as jointly dependent variables. The method we suggest circumvents this deficiency and is more consistent with the notion of qualification discrimination.

Significantly, there is a stark contrast between our estimates of qualification discrimination and previous estimates—rather than finding no qualification discrimination, we find significant discrimination against men. Thus, although men are overpaid according to traditional wage regressions, they are overqualified according to our improved qualification regressions. Such sharp contradictions in empirical measures should heighten our sensitivity both to the issue of what is meant by employment discrimination and to the methods employed to measure it.

The remainder of the paper is organized as follows. Section II briefly discusses the

deficiencies of previous methods of measuring qualification discrimination and offers an alternative method that avoids these deficiencies. Section III presents estimates of qualification discrimination for men and women based on the alternative method. Traditional wage regression estimates of discrimination are also presented for comparison. A discussion of the implications of our findings is presented in section IV. A concluding section summarizes our major points.

II. A Procedure for Qualification Regressions

The multidimensional nature of employment qualifications is an obvious impediment to the implementation of qualification regression, since standard regression techniques cannot be used with multiple dependent variables. To some extent, this is also a problem for wage regressions, although wages dominate total compensation to a greater degree than any one qualification is likely to dominate total qualifications.

Kamalich and Polachek [8] offer one approach: generate a series of regressions, one for each qualification, by regressing each qualification in turn on wages and the remaining qualifications. There are at least two serious objections to this approach. First, most of the qualifications are still included along with sex (or race) as explanatory variables in any one regression. This treatment of the remaining qualifications is inconsistent with the point of reverse regression, which is to measure discrimination with only wages and sex (or race) as explanatory variables.

Second, the series of regressions are almost certain to yield a jumble of conflicting results. Every relative qualification must turn in the same direction to support a definite conclusion regarding qualification discrimination. Kamalich and Polachek, for example, conclude that for the economy as a whole clear-cut qualification discrimination does not exist because there is evidence of both under and overqualification for women and blacks. They are left with this conclusion because their approach provides no way of reconciling multiple qualifications.

Conway and Roberts [2] do offer an approach that deals directly with multiple qualifications. To obtain a single index of qualifications for use as a dependent variable, they weight each qualification by the corresponding coefficient in a wage regression. This index is then regressed on wages, sex (or race), and other variables. Although the Conway-Roberts procedure does deal with multiple qualifications, it does so by resorting to the results of wage regression—a regression in which *qualifications*, rather than *wages*, are held constant. This also violates the point of qualification regression, which is to compare qualifications holding wages constant.

We propose a similar two-step procedure, but one in which the weights used to collapse multiple qualifications into a single index are obtained from an empirical specification that holds wages constant and treats qualifications simultaneously as jointly dependent variables. Canonical correlation exhibits just these properties. That is, canonical correlation finds the linear combination of dependent variables and the linear combination of independent variables that maximize the correlation between the two linear combinations (the canonical variates). In this way weights can be obtained for the various qualifications using only wages and sex as explanatory variables. The resulting index of qualifications can then be regressed on wages and sex.

Table I. Variable Means by Sex and Race (PSID 1980)

	Blue Collar				White Collar			
Variables	White Males	White Females	Black Males	Black Females	White Males	White Females	Black Males	Black Females
LOGWAGE	6.57	6.09	6.32	5.99	6.92	6.36	6.57	6.23
PROF					.45	.31	.28	.24
MANAG					.38	.12	.24	.08
FORE	.52	.06	.26	.02				
ED	11.43	11.45	10.52	10.88	14.60	13.44	13.15	12.77
EDSQ	134.68	134.76	117.89	121.91	218.10	185.33	177.38	166.71
COLDEG	.03	.04	.02	.01	.50	.28	.18	.16
ADVDEG	.00	.00	.00	.00	.16	.06	.04	.03
TEN	6.19	4.33	6.22	5.41	6.40	4.29	4.68	3.71
TENSQ	91.18	45.89	90.42	63.56	98.24	41.56	49.81	28.61
EX	16.84	11.84	16.10	13.74	18.23	11.94	13.46	10.86
EXSQ	417.78	231.36	386.18	301.46	442.94	206.17	258.24	185.09
Sample size	986	287	704	373	813	763	152	333

III. Comparisons of Wage and Qualification Discrimination

Data and Empirical Specifications

The data used in our analysis are from the 13th wave (1980) of the Panel Study of Income Dynamics (PSID). Approximately 5000 households are included in the survey, but our sample is restricted to individuals between 18 and 62 years of age who were earning at least one dollar per hour in current, full-time employment, and for whom there is a complete report of all variables used in the analysis. These restrictions lead to a sample of 4401 individuals.

Due to concerns regarding structural differences, separate analyses are performed for blue-collar and white-collar workers, segregated by race. Our empirical specifications are taken from familiar specifications in the wage and discrimination literature [7; 9; 10; 11]. Table I presents means of the standard human capital variables used in our primary regressions. *LOGWAGE* is the logarithm of the hourly wage. *PROF*, *MANAG*, and *FORE* are binary dummy variables, respectively, for professional-technical workers, managers, and foremen and skilled craftsmen. Among white-collar workers sales and clerical workers are the reference group for *PROF* and *MANAG*. For blue-collar workers, semi-skilled and unskilled workers are the reference group for *FORE*. *ED* and *EDSQ* are completed years of schooling and *ED* squared, respectively. *COLDEG* (college degree) and *ADVDEG* (advanced degree) are included in addition to *ED* and *EDSQ* to account for additional nonlinearities and degree effects of education. *TEN* and *TENSQ* represent tenure on the current job in years and *TEN* squared, respectively. *EX* and *EXSQ* represent the total

Table II. Wage and Qualification Regressions for Blue-Collar Workers (PSID 1980)

	Whites		Blacks	
Variables	Wage Regression	Qualification Regression	Wage Regression	Qualification Regression
Intercept	5.38 (32.24)	−6.71 (−19.96)	5.42 (45.25)	−6.31 (−14.80)
LOGWAGE		1.00 (18.28)		1.00 (14.12)
MALE	.33 (13.24)	.32 (5.59)	.28 (12.06)	.17 (2.80)
FORE	.20 (9.25)	1.26	.18 (6.25)	1.68
ED	.035 (1.12)	.05	−.003 (−.12)	−.25
EDSQ	.0007 (.47)	.005	.003 (2.38)	.026
COLDEG	−.103 (−1.31)	−.57	−.173 (−1.78)	−.84
ADVDEG	.412 (1.19)	1.49	−.323 (−1.26)	−1.40
TEN	.02 (5.15)	.02	.01 (2.58)	.02
TENSQ	−.0004 (−3.00)	−.0010	−.0003 (−1.64)	.0001
EX	.017 (5.00)	.07	.017 (4.70)	.11
EXSQ	−.00028 (−3.57)	−.0012	−.00023 (−2.65)	−.0017
R^2	.39	.32	.30	.21
Number	1263	1263	1077	1077

Notes: Coefficients are followed by *t*-statistics in parentheses. Those coefficients without *t*-statistics are rescaled canonical weights used to construct the qualification index for that regression. See text for explanation.

number of years of labor market experience since age 18 and *EX* squared, respectively. As expected, the largest differences between men and women are in current job tenure (*TEN*) and in total experience (*EX*).

For comparison, both wage and qualification regressions are performed for each group of workers. In the wage regressions, the logarithms of hourly wage (*LOGWAGE*) is regressed on years of education (*ED*), years of education squared (*EDSQ*), college degree (*COLDEG*), advanced degree (*ADVDEG*), tenure (*TEN*), tenure squared (*TENSQ*), experience (*EX*), experience squared (*EXSQ*), dummy variables for occupation group, and a dummy variable for males (*MALE*). Except for the inclusion of degree variables, this specification is similar to those of Kamalich and Polachek.

Table III. Wage and Qualification Regressions for White-Collar Workers (PSID 1980)

	Whites		Blacks	
Variables	Wage Regression	Qualification Regression	Wage Regression	Qualification Regression
Intercept	5.82 (31.49)	−6.83 (−25.53)	5.26 (10.37)	−6.37 (−11.40)
LOGWAGE		1.00 (23.88)		1.00 (11.18)
MALE	.33 (16.00)	.35 (8.30)	.27 (7.67)	.12 (1.48)
PROF	.15 (5.76)	.71	.03 (.77)	.21
MANAG	.19 (7.13)	1.06	.12 (2.37)	.98
ED	−.021 (−.70)	−.14	.048 (.59)	−.07
EDSQ	.0023 (1.84)	.011	.00059 (.18)	.015
COLDEG	.027 (.69)	−.049	−.026 (−.34)	−.462
ADVDEG	.072 (2.07)	.171	.087 (.86)	.231
TEN	.036 (.86)	−.01	.005 (.49)	−.01
TENSQ	−.0144 (−.51)	.0004	−.0001 (−.40)	.002
EX	.034 (9.29)	.10	.033 (5.48)	.16
EXSQ	−.0005 (−6.00)	−.0013	−.0007 (−4.74)	−.004
R^2	.50	.44	.34	.26
Number	1576	1576	485	485

Notes: Coefficients are followed by *t*-statistics in parentheses. Those coefficients without *t*-statistics are rescaled canonical weights used to construct the qualification index for that regression. See text for explanation.

The qualification regressions are performed in the two-step method described in the previous section. The qualification index is derived from canonical correlation estimates in which the left-hand-side variables are all those independent variables in the wage regressions except *MALE*. The right-hand-side variables are *LOGWAGE* and *MALE*. In the second step, the qualification index is regressed on the right-hand-side variables *LOGWAGE* and *MALE*. So that comparisons can be made between the wage and qualification regressions, the parameter estimates in the qualification regressions are measured in (log) wage units by normalizing the parameters on *LOGWAGE*. This implies that the coefficient on *LOGWAGE* is exactly one.

Estimates

Tables II and III report the results of the regressions for blue-collar and white-collar workers, respectively. In every case, contradictory evidence of discrimination is provided by the wage and qualification regressions. For blue-collar workers, black men have a wage advantage of .28 (32 percent)[1] over black women with the same qualifications, but are overqualified by .17 (19 percent) relative to black women earning the same wage. White men have a wage advantage of .33 (39 percent) over white women with the same qualifications, but are overqualified by .32 (38 percent) relative to white women earning the same wage. As expected, the foreman dummy (*FORE*) is positive and significant in both wage regressions. The directions of the effects of the qualification variables are the same in the wage and qualification regressions (except for *TENSQ* for blacks, which is not significant in the wage regression), but the magnitudes of the effects are often substantially different (no formal tests are possible). Therefore, the use of coefficients from wage regressions as weights for a qualification index appears to be inappropriate.

The results in Table III for white-collar workers are similar to those for the blue-collar workers in Table II. Black men have a wage advantage of .27 (31 percent) over equally-qualified black women, but are overqualified by .12 (13 percent) relative to black women with the same wage. The latter result, however, is not statistically significant (at the .05 level). White men have a wage advantage of .33 (39 percent) over white women, but are overqualified by .35 (42 percent) compared to white women earning the same wage. As expected, the *PROF* and *MANAG* coefficients are generally positive in both the wage and qualification regressions. The wage and qualification regressions for both blue-collar and white-collar workers have roughly the same explanatory power.

Sensitivity

We should also note that the substance of our empirical conclusions is not altered by considering a variety of alternative specifications. These include separate regressions for 1) separate occupations (those occupations denoted by dummy variables in the previous specifications); 2) union and nonunion workers; 3) private and governmental workers; and 4) workers above and below the age of 35. In addition our conclusions are not sensitive to the introduction of interactions between male and logwage in the qualification regression. Results for the latter, however, do indicate that low-wage males are relatively more overqualified than high-wage males.

IV. Remarks

Discovery of such sharp contradictions in our measure of discrimination begs for reconciliation. Is one measure of discrimination superior to the other? Do the two types of discrimination somehow "cancel out" so that men receive no net advantage?

Regrettably, an easy reconciliation is not forthcoming. As we discussed earlier, in the absence of any sources of error in the relationships among wages and productivity, wage and qualification discrimination are necessarily the same. However, these errors are far from absent. Examples of possible sources of errors abound. For instance, hourly wages

1. The percentage effect on dummy variables in semilogarithmic equations is discussed in Halverson and Palmquist [6] and is equal to 100 times $[\exp(c) - 1]$, where c is the coefficient on the dummy variable.

and years of schooling are imperfect proxies for total compensation and for the productive capacity produced by education. Experience, tenure, and education are all possible sources of nonrandom errors-in-variables that could be related to sex. Possible omitted variables include some measure of job or employee quality, which well may be correlated with sex. The magnitude and direction of biases introduced into both qualification and direct regressions by these errors are generally indeterminate.

One possible employment situation that might yield the situation we discover — males receiving higher pay than females of equal qualifications but males being better qualified than females earning the same amount — is where women recently hired have fewer qualifications than men with the same salary, while older female incumbents receive a wage lower than their male peers. In this situation, we might expect to observe qualification discrimination against men in earlier career stages, but wage discrimination against women in later career stages, women whose work lives began prior to the passage of Title VII of the Civil Rights Act. However, there is no difference in our findings between workers above and below the age of 35. In both cases, we have evidence that men are overqualified relative to women earning the same wage, and overpaid relative to women with the same qualifications.

Another possibility is that men and women on average are in occupations with noncomparable characteristics, especially noncomparable experience profiles. Since women typically have flatter profiles than men, mixing occupations with distinctly different profiles could yield the anomalous result we find. This possibility is consistent with evidence that male overqualification decreases with the level of the wage. Again, however, separate analyses by occupation begs the question of occupation segregation.

The implication of our results for public policy is critical. Public policy to reduce employment discrimination will be most effective if there are valid methods of identifying the existence and extent of discrimination, and of verifying if policies directed at countering discrimination have achieved their goal. The fact that two seemingly valid methods of discrimination provide directly contrasting evidence raises the danger that evidence of discrimination, or of nondiscrimination, can be selected in accordance with prejudices. Yet it is inappropriate to assume that the two types of discrimination somehow cancel out, so that males receive no net advantage. Instead, such sharp contradictions in empirical measures should underscore the difficulty of using aggregate measures as evidence of discrimination.

V. Conclusion

Evidence of discrimination has been provided traditionally by examining differences in wages for individuals with equivalent (observed) job qualifications. Recently some researchers have attempted to measure discrimination by differences in qualifications for individuals with the same wage. The empirical methods employed, however, have been inappropriate. In this paper we suggest an appropriate method of estimating qualification discrimination that relies on canonical correlation. Using data from the 1980 Panel Study of Income Dynamics, and after controlling for differences in expected productivity, we find that men have a wage advantage over women with equivalent qualifications ranging from 31 percent (for white-collar blacks) to 39 percent (for both blue-collar and white-collar whites). Al-

though this would appear to provide unambiguous evidence of extensive discrimination against women, the same data yield evidence of discrimination against men which is nearly as severe. Men are overqualified relative to women receiving the same wage, with the degree of overqualification ranging from 13 percent for white-collar blacks to 42 percent for white-collar whites.

Joni Hersch
Joe A. Stone
University of Oregon
Eugene, Oregon

References

1. Birnbaum, Michael H. "Procedure for the Detection and Correction of Salary Inequities," in *Salary Equity*, edited by Thomas R. Pezzullo and Barbara E. Brittingham. Lexington, Mass.: Lexington Books, 1979, pp. 121–44.

2. Conway, Delores and Harry Roberts, "Reverse Regression, Fairness and Employment Discrimination." *Journal of Business and Economic Statistics*, January 1983, 75–85.

3. _____, "Rejoinder to Comments on 'Reverse Regression, Fairness, and Employment Discrimination'." *Journal of Business and Economic Statistics*, April 1984, 126–39.

4. Ferber, Marianne A. and Carole A. Green, "What Kind of Fairness is Fair? A Comment on Conway and Roberts." *Journal of Business and Economic Statistics*, April 1984, 111–13.

5. Goldberger, Arthur S., "Redirecting Reverse Regression." *Journal of Business and Economic Statistics*, April 1984, 114–16.

6. Halverson, Robert and Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review*, June 1980, 474–75.

7. Hanoch, Giora, "An Economic Analysis of Earnings and Schooling." *Journal of Human Resources*, Summer 1967, 310–29.

8. Kamalich, Richard and Solomon Polachek, "Discrimination: Fact or Fiction? An Examination Using an Alternative Approach." *Southern Economic Journal*, October 1982, 450–65.

9. Malkiel, Burton G. and Judith Malkiel, "Male-Female Pay Differentials in Professional Employment." *American Economic Review*, September 1973, 693–705.

10. Mincer, Jacob. *Schooling, Experience and Earnings*. New York: Columbia University Press, 1974.

11. Oaxaca, Ronald, "Male-Female Differentials in Urban Labor Markets." *International Economic Review*, October 1973, 693–709.

12. Polachek, Solomon W., "Potential Biases in Measuring Male-Female Discrimination." *Journal of Human Resources*, Spring 1975.

13. Theil, Henri. *Principles of Econometrics*. New York: John Wiley & Sons, Inc., 1971.