University of Wisconsin-Madison

IRP Discussion Papers

Arthur S. Goldberger

REVERSE REGRESSION AND SALARY DISCRIMINATION

DP # 743-84

Institute for Research on Poverty
Discussion Paper 743-84

Reverse Regression and Salary Discrimination

Arthur S. Goldberger

January 1984

Presented at the winter meetings of the Econometric Society, San Francisco, December 28-30, 1983. For research support I am grateful to the William F. Vilas Trust and to the Center for Advanced Study in the Behavioral Sciences. For instructive comments, I am grateful to John Abowd, Michael Birnbaum, Gail Blattenberger, Glen Cain, Gary Chamberlain, Delores Conway, Christopher Flinn, Edward Leamer, Thomas MaCurdy, Charles Manski, Jon Peck, John Pencavel, and Solomon Polachek. Also issued as Social Systems Research Institute Workshop Series no. 8401, University of Wisconsin-Madison.

ABSTRACT

Reverse regression has recently been proposed to assess salary discrimination by gender or race. We consider several stochastic models and find the one that justifies reverse regression. Testable implications are deduced, and the analysis is illustrated with empirical material.

REVERSE REGRESSION AND SALARY DISCRIMINATION

by Arthur S. Goldberger

1. Introduction

Are men paid more than equally-productive women? The conventional approach to answering this question is to regress an earnings variable (y) upon a set of productivity-related "qualifications" ($\underline{x} = (x_1, \ldots, x_k)'$) and a gender dummy (z: z=1 for men, z=0 for women). This estimates

(1) $$E(y|\underline{x},z) = \underline{b}'\underline{x} + a\ z,$$

in which the coefficient a is taken to be the discriminatory premium paid to men. I have introduced, and will maintain, several conventions and assumptions: all regressions are linear-additive, all variables have zero means for women, and sampling variability is neglected. The raw difference between men's mean y and women's mean y is denoted m: we write $E(y|z) = mz$.

Over the past decade, this direct approach has been used on national samples (e.g. Oaxaca, 1973), on individual firm data (e.g. Malkiel & Malkiel, 1973), and in various discrimination suits: see Finkelstein (1980), Baldus & Cole (1980: Chapter 8; 1982: Chapter 8), Bloom & Killingsworth (1982). The usual finding, namely $a > 0$, is offered as evidence of salary discrimination in favor of men, that is against women: among men and women possessing equal qualifications, men are paid more. Similarly for the analogous white-black comparison, for which I adopt the analogous coding: z = 1 for whites, z = 0 for blacks.

Recently, an alternative approach has been proposed, one which turns the question around to ask "Are men less qualified than equally-paid women?": Roberts (1980), Birnbaum (1979a,b; 1981), Kapsalis (1982), Dempster (1982), Kamalich & Polachek (1982), Conway & Roberts (1983). The version of the new approach which particularly concerns me is one that is intended to handle the multiple qualification case: Let $q = \underline{b}'\underline{x}$ denote the scalar index of qualifications implied by the direct regression. Now regress q upon y and z, estimating

$$(2) \qquad E(q\,|\,y,z) = c\ y + d\ z,$$

and take the coefficient d to be the excess qualifications required of men. On this approach, the finding $d < 0$ is needed to establish discrimination in favor of men: among men and women receiving equal salaries, the men possess lower qualifications.

One might anticipate that this new, reverse, approach would give the same qualitative answer as the customary, direct, approach. If men are paid more than equally-qualified women, then they are less qualified than equally-paid women: $a > 0 \Rightarrow d < 0$. While that reasoning applies to a deterministic relationship, where $y = \underline{b}'\underline{x} + az = q + az$ implies $q = y - az$, it is by no means guaranteed empirically where relationships are far from deterministic. Hashimoto & Kochin (1980) provide a striking example by tabulating median income vs. education, and vice-versa, for whites and non-whites using the 1960 Census. As shown in Table 1, at each level of education, average income was higher for whites ($a > 0$), but at each income level, average education was also higher for whites ($d > 0$). They describe this as a "riddle" to be explained as "an artifact of errors in variables".

Birnbaum (1979b) reanalyzes a study of 1976 salaries for 119 male and 153 female faculty members at the University of Illinois matched by department. He reports: "On the average males are paid about $2,000 more than females with the same number of publications" while "females publish about 2 fewer articles per five years than males who receive the same salary". Thus, both a and d are positive, although "one would expect women in a discriminatory situation to have published more than men with the same salaries". He rationalizes this puzzle, or paradox, by supposing that salary, publications, and other measured qualifications are fallible measures of "quality" (i.e. true productivity). He also suggests that reverse regression should be used along with direct regression to assess discrimination.

Kamalich & Polachek (1982) report results for gender as well as race, using 4542 observations in the 1976 Panel Survey of Income Dynamics. As Table 2 shows, their direct regressions (of log wages upon schooling, tenure, experience, and the group dummy) indicate substantial discrimination in favor of men and of whites ($a > 0$). But the system of reverse regressions (of each qualification variable upon log wages and the group dummy), also shown in the table, is indicative in some cases of discrimination in favor of women and of blacks ($d > 0$). After further elaboration of the new approach, the authors conclude: "for the economy as a whole <u>clear-cut discrimination</u> [in favor of men and of whites] <u>does not</u> exist. ... In fact, ... there is evidence of reverse discrimination [in favor of blacks]." They motivate the use of reverse regression by asserting that "productivity proxies mismeasure true productivity."

Conway & Roberts (1983), working with data for 274 employees of a Chicago bank in 1976, report a direct regression of log salary upon six educational, experience, and age variables, along with gender. The value of a is .148 (standard error .036) indicating that men are overpaid by about 16%: see Table 3. Their reverse regression of $q = \underline{b}'\underline{x}$ upon log salary and gender, however, shows d = -.0097 (standard error .0202). "Hence", they conclude, "in this application, direct regression shows a substantial and significant female salary shortfall for given qualifications, and a near standoff of qualifications for given salary." These authors motivate reverse regression, in part, by pointing out that:

> "In regression studies of discrimination, not all pertinent job qualifications are available to the statistician. Indeed, the job qualifications actually available typically comprise a very incomplete listing of pertinent qualifications for any job... The problem of omitted job qualifications points to the weakness of a direct-regression-adjusted income differential as a definition of discrimination."

Abowd, Abowd, & Killingsworth (1983) work with a large sample from the 1976 Survey of Income and Education. Table 4 shows a portion of their results, comparing whites with several ethnic groups in turn. The y variable is log wage, $\underline{x}$ contains about thirty educational, age, experience, and locational variables, and $q = \underline{b}'\underline{x}$ is the dependent variable in the reverse regression. Observe that in each case, the direct coefficient a is positive (indicating whites are favored), while the reverse coefficient d is also positive (indicating that whites are disfavored). These authors, who are by no means advocates of reverse regression, introduce it as a procedure that may correct for the measurement error bias which is associated with observed qualifications being imperfect measures of true productivity.

At this point, it seems fair to summarize the empirical results that we have been seeing as follows: Reverse regression points to a lower estimate of salary discrimination (in favor of men, or of whites) than does direct regression; indeed it often suggests reverse discrimination (against men, or against whites). If so, the new approach has obvious attractions for defendants (employers) in discrimination suits, and indeed has been already used in that context. It also has attractions for academic researchers who seek dramatic and counter-intuitive results. But the scientific case for reverse regression, or rather for choice of regression, will properly rest on a stochastic model of salary determination. As we have seen, advocates of reverse regression have attempted to make such a case by referring to the fact, or presumption, that the qualification variables in $\underline{x}$ do not exhaust (i.e., are merely proxies for) the productivity assessment actually used by the employer in setting salaries.

My concern in this paper is to evaluate the statistical argument for reverse regression. Section 2 sketches the errors-in-variable argument in the single-qualification case. In Section 3, I evaluate a series of claims about direct and reverse regression that have appeared in the literature. In Section 4, I turn to the multiple-qualification case, specifying several models of salary determination, and evaluate the validity of estimators. Section 5 re-assesses several of the empirical studies from that perspective. Section 6 provides concluding remarks.

Some further notation is in order. When each qualification is taken in turn as the dependent variable, we get a system of reserve regressions:

$$E(x_j | y,z) = c_j\, y + d_j\, z, \qquad (j=1,\ldots,k) \tag{3}$$

which may also be written in multivariate format as

(4) $E(\underline{x}|y,z) = \underline{c}\, y + \underline{d}\, z,$

where $\underline{c} = (c_1,\ldots,c_k)'$ and $\underline{d} = (d_1,\ldots,d_k)'$. Also, it is convenient to make the direct and reverse regression coefficients more readily comparable. To do this, rewrite the reverse regression to put y on the left-hand side, thus expressing the qualifications differential in units of y. While users of reverse regression are not always clear on this point, I believe it is fair to say that they would take, as the alternative to the direct estimate (a), the gender coefficients

(5) $a_j^* = -d_j/c_j \qquad (j=1,\ldots,k)$

when individual reverse regressions (3) are run, and correspondingly take

(6) $a^* = -d/c$

when the composite reverse regression (2) is run. On this understanding, the naive anticipation was that a and a* would have the same sign, and what we have been seeing is that this anticipation is not always realized. To fix ideas, the Conway & Roberts results (see Table 2) would be reported as

$$a = .148, \quad a^* = -(-.0097)/.316 = .031:$$

direct regression shows men overpaid by about 16%, while reverse regression shows the same men overpaid by only about 3%.

2. Errors in Variables

It is plausible that the employer, in setting y, had access to more productivity-relevant information than is contained in $\underline{x}$, the vector of measured qualifications that is available to the statistician: see Roberts (1980), Dempster (1982). If so, and if that missing information is correlated with gender, then it is possible that in the direct regression the gender variable z will be serving in part as a proxy for the omitted variables. Consequently, the direct estimate a may be spurious, and the reverse estimate a* may be preferable.

For the situation in which $\underline{x}$ contains a single variable, several authors have made this case against direct regression in terms of a classical errors-in-variables model. Our version, which is formulated to facilitate extension to the multiple-x situation, runs as follows:

$$\text{(7a)} \qquad y = p + \alpha z + v$$

$$\text{(7b)} \qquad p = \beta x^*$$

$$\text{(7c)} \qquad x^* = \mu z + u$$

$$\text{(7d)} \qquad x = x^* + \varepsilon,$$

with y = salary, p = productivity, z = gender, x* = true qualification, and x = measured qualification. We take v, u, ε to be mutually independent with expectations zero, variances σ_v^2, σ_u^2, and σ_ε^2, all independent of z. In (7a) salary is a stochastic function of productivity and of gender. The structural parameter of interest is α, the discriminatory premium paid to men. In (7b) productivity is an exact function of true qualification; we take $\beta > 0$. In (7c) the expectation of true qualification is allowed to differ by gender.

We will suppose that $\mu > 0$ since that is considered to be the empirically relevant case. Finally, in (7d), measured qualification is a fallible indicator of true qualification, in the classical errors-in-variable sense. Figure 1A gives a path-diagram representation of the model.

Now consider the direct regression of y upon x and z: $E(y|x,z) = bx + az$. Since variances and covariances are the same for both genders, we can calculate the common slope as

$$b = C(x,y|z)/V(x|z) = C(x^*,p|z)/V(x|z) = \beta V(x^*\ z)/V(x|z) =$$

(8) $$b = \beta\pi^*,$$

say, where

(9) $$\pi^* = V(x^*|z)/V(x|z) = \sigma_u^2/(\sigma_u^2 + \sigma_\varepsilon^2)$$

lies in the unit interval. And then the gender coefficient follows as

$$a = E(y|z=1) - b\ E(x|z=1) = \alpha + E(p|z=1) - b\ E(x^*|z=1)$$
$$= \alpha + \beta\mu - b\mu = \alpha + (\beta - b)\mu =$$

(10) $$a = \alpha + (1-\pi^*)\beta\ \mu.$$

Evidently (with $0 \leq \pi^* < 1$, $\beta > 0$, $\mu > 0$) the direct regression estimator of α is biased upwards: Hashimoto & Kochin (1980), Roberts (1980), Birnbaum (1979a, 1981), Robbins & Levin (1982), Abowd, Abowd, & Killingsworth (1983). Since x is a fallible measure of x^* ($\pi^* < 1$) and z is a positive correlate of x^* ($\mu > 0$), the coefficient on z is positively contaminated. As with most interesting aspects of errors-in-variable models, this conclusion has its parallel in permanent-income theory: see Friedman (1953: 79-90) on black-white consumption functions.

Now consider the reverse regression of x upon y and z: $E(x|y,z) = cy + dz$. We calculate

$$c = C(x,y|z)/V(y|z) = V(x^*|z)/(\beta^2 V(x^*|z) + V(v|z)) = \beta\sigma_u^2/(\beta^2\sigma_u^2 + \sigma_v^2) =$$

(11) $$c = \pi/\beta,$$

say, where

(12) $$\pi = V(p|z)/V(y|z) = \beta^2\sigma_u^2/(\beta^2\sigma_u^2 + \sigma_v^2)$$

lies in the unit interval. And then the gender coefficient follows as

$$d = E(x|z=1) - c\,E(y|z=1) = \mu - c(\alpha+\beta\mu) = -c\alpha + (1-\beta c)\mu =$$

(13) $$d = -c\alpha + (1-\pi)\mu.$$

To obtain the implied estimator of the discrimination parameter α, we rearrange the reverse regression to put y on the left-hand side, and find

(14) $$a^* = -d/c = \alpha - (1-\pi)\mu/c = \alpha - \left(\frac{1-\pi}{\pi}\right)\beta\mu.$$

Evidently (with $0 \leq \pi < 1$, $\beta > 0$, $\mu > 0$), this reverse regression estimator of α is biased downward: see Roberts (1980), Birnbaum (1981), Abowd et al. (1983), Solon (1983).

At this stage of the argument, reverse regression merely provides a lower bound to α. But some proponents of reverse regression push the argument a step further by taking the salary function to be deterministic. Suppose then that v = 0, so $\sigma_v^2 = 0$, so $\pi = 1$. Then from (11)-(14), $c = 1/\beta$, $d = -c\alpha$, and $a^* = \alpha$: reverse regression gives an unbiased estimator of α. Indeed, with v = 0, the conclusion is evident by rearranging (7a-b) to get $y = \beta x^* + \alpha z$, whence $x^* = (1/\beta)y - (\alpha/\beta)z$, and

$$(15) \qquad x = (1/\beta)y - (\alpha/\beta)z + \varepsilon.$$

With ε independent of z and y, we have

$$(16) \qquad E(x|y,z) = (1/\beta)y - (\alpha/\beta)z,$$

which will indeed be correctly estimated by regression of x upon y and z. For the analogous discussion in permanent-income theory, see Friedman (1957: 200-206).

The rationales offered for specifying $\sigma_v^2 = 0$ have been rather casual: Roberts's numerical example (1980: 183-186) just takes it for granted; Dempster (1982: 12) is "somewhat skeptical about the existence of a chance mechanism whereby the employer creates a random disturbance and adds it" to his best assessment of productivity; while Kamalich & Polachek (1982: 453-456) simply reproduce Roberts's numerical example.

On my reading of the literature, the simple model of this section has served as the underpinning for sweeping generalizations about the defects of direct, and the virtues of reverse, regression. To anticipate the pitfalls of that mode of argument, readers may want to consider two questions:

(i) granted that measured qualifications give only incomplete information on the employer's productivity assessment, does it follow that they are fallible measures in the errors-in-variable sense?

(ii) granted that measured qualifications <u>are</u> fallible measures, are they fallible measures of true productivity, or of its determinants?

3. <u>Claims and Impressions</u>

As far as I can see, much of the recent literature on reverse regression assessment of discrimination relies heavily, if not always explicitly, on the univariate errors-in-variable specification of the previous section. Critics of direct regression and proponents of reverse regression have made various claims which I will assemble below. In fairness, I must remark that I may have taken some of the quotations out of context, and a very close reading of the articles may reveal that the authors' claims were sufficiently qualified as to be justified. But I do believe that most readers of these articles will have come away with the impressions that the assertions made hold quite generally.

To avoid repetition in what follows, let us take it for granted that "discrimination" means $\alpha > 0$, and that men (or whites) rate higher than women (or blacks) on all productivity variables (including true productivity).

Here is my list, along with illustrative citations.

3.1. <u>The direct regression estimate of discrimination is biased</u> (upward) <u>unless measured qualifications fully capture productivity</u>.

Wolins (1978: 717). "Variables such as number of publications are, however, fallible indicators of constructs, and being fallible they control incompletely for the target construct, research productivity... Covariance analysis [i.e., direct regression]... is known to be biased... The group higher on the fallible covariate will tend to appear disproportionately higher on the variate...even when there would be no such disproportionate difference if an infallible covariate were used."

McCabe (1980:213). "If there are merit variables, positively associated with salary, and the protected group means are less than the unprotected group means on these variables, then an overestimate of the salary differential will be obtained by a linear analysis whenever these variables are excluded from the analysis."

Roberts (1980: 177). "There is good reason to expect...that the omission of variables...may have a biasing effect, tending to give the appearance of discrimination when none exists. Moreover the danger of underadjustment...can be expected to affect almost all statistical studies of possible discrimination."

Roberts (1980: 186, 188). "It is a consequence of the fact that statisticians must work with (crude) proxies rather than true productivity... Underadjustment...was due to the fact that the variable proxy can be thought of as an imperfect measurement of true productivity."

Humphreys (1981: 1192-1193). "One must assume that the correlation between measured merit and the latent trait is equal to unity if one intends to consider only the [direct regression estimate]... for either theoretical discussion or social action."

Kamalich & Polachek (1982: 453, 454, 460). "Estimates of discrimination (the race and sex coefficients) are biased when productivity proxies mismeasure true productivity... Any regression of wage on productivity proxies in which one group tends to have higher productivity will run into this type of bias... We have shown that the traditional method of examining discrimination...is clearly biased. Failure to account for measurement error in productivity proxies tends to overestimate discrimination."

3.2. <u>When measured qualifications do not fully capture productivity, the reverse regression estimate of discrimination is unbiased.</u>

Roberts (1980: 177, 186-187). "Reverse regression can cope with this bias... The statistician need merely compare mean values of the proxy between males and females at each given salary level."

Kapsalis (1982: 272). "There is no reason to expect that the new measure is downward biased... There is no reason to expect that the new measure is biased."

Kamalich & Polachek (1982: 456). Having reproduced Roberts's numerical illustration (a univariate errors-in-variable model), these authors turn to the multivariate case, and write "From this illustration, it should be clear that the appropriate reverse regression consists of a formulation in which each productivity proxy is a dependent variable, and sex (race), wage, and [the other productivity proxies]... serve as independent regressors." They go on to suggest that dropping the other proxies is also acceptable: "This simplified version has the advantage of minimizing errors of measurement problems, though it may suffer from problems of omitted variables. Further this simple model serves as an upper bound for discrimination". More cryptic still is their remark (p. 460) that "biases can creep in" to reverse regression via simultaneity and multicollinearity.

3.3. <u>If discrimination is present then both the direct and the reverse regression estimates</u> (i.e., a and a*) <u>must be positive.</u>

Birnbaum (1979b: 719). "In order to demonstrate systematic sex discrimination, it must be shown not only that women earn less on the average than men of the same qualifications, but also that they are more qualified on the average than men receiving the same salary."

Kamalich & Polacheck (1982: 450). "If discrimination exists, one would expect to find blacks and women to have <u>higher mean qualifications for any given wage level</u>." Having run the system of reverse regressions (our (3)) and found that the signs of the d_j (hence of the a_j^*) are mixed, they say that "the pattern of mixed positive and negative coefficients...is consistent with nondiscrimination, as shortfalls in one area for particular groups are offset by strengths in other proxies." (p. 459).

3.4. <u>The direct and reverse regression estimates provide bounds for the true discrimination parameter</u>.

Abowd et al. (1983: 9). "The importance of direct and reverse regression analyses of wage differentials, then, is simply that in the presence of measurement error in both p and y, the two procedures will produce an upper and a lower bound for the actual magnitude of discrimination."

3.5. <u>Reverse regression is more direct than direct regression.</u>

Kamalich & Polachek (1982: 461). It "measures discrimination <u>directly</u>, and not indirectly as a residual, as done in all past [i.e., direct] analyses."

Many of these claims can be disposed of immediately once we recognize that the errors-in-variable specification (7) is not the only one which permits imperfect correlation between measured qualifications and productivity. (See also Weisberg & Tomberlin (1983: 399-400)). Suppose that

(17a) $y = p + \alpha z + v$

(17b) $p = \beta x + \varepsilon$

(17c) $x = \mu z + u.$

We take v, ε, u to be mutually independent with expectations zero, variances σ_v^2, σ_ε^2, and σ_u^2, all independent of z. In (17a) salary is a stochastic function of productivity and of gender; the structural parameter of interest is still α. In (17b) productivity is a stochastic function of measured qualification; we take $\beta > 0$. In (17c) the expectation of measured qualification is allowed to differ by gender; we suppose $\mu > 0$. Figure 1B gives the path diagram.

Now consider the direct regression of y upon x and z: $E(y|x,z) = bx + az$. Since variances and covariances are independent of gender, we can calculate the common slope as

(18) $b = C(x,y|z)/V(x|z) = C(x,p|z)/V(x|z) = \beta V(x|z)/V(x|z) = \beta,$

and then the gender coefficient follows as

(19) $a = E(y|z=1) - b\ E(x|z=1) = \alpha + \beta\mu - \beta\mu = \alpha.$

Clearly the direct regression estimator of α is unbiased, <u>even though</u> x

<u>is not a perfect correlate of p</u>. From (17b) we see that the unconditional squared correlation between x and p is

$$(20) \qquad \rho^2 = \beta^2 V(x)/(\beta^2 V(x) + \sigma_\varepsilon^2).$$

Observe that x can be a proxy for p (in the imperfect correlate sense) without being a fallible measure of p (in the classical errors-in-variable sense). Confusion on this elementary distinction has prevailed in the recent literature.

Our conclusion on the unbiasedness of direct regression can be established more simply: substitute (17b) into (17a) to get

$$(21) \qquad y = \beta x + \alpha z + (\varepsilon + v) = \beta x + \alpha z + t,$$

say. With $t = \varepsilon + v$ independent of x and z, we get

$$(22) \qquad E(y|x,z) = \beta\ x + \alpha\ z,$$

which is indeed unbiasedly estimated by direct regression.

Consider instead the reverse regression of x upon y and z: $E(x|y,z) = cy + dz$. We calculate

$$(23) \qquad c = C(x,y|z)/V(y|z) = \beta\ \sigma_u^2/(\beta^2\sigma_u^2 + \sigma_t^2) = \pi/\beta,$$

say, where

$$(24) \qquad \pi = \beta^2\sigma_u^2/(\beta^2\sigma_u^2 + \sigma_t^2)$$

lies in the unit interval. And then the gender coefficient is

$$(25) \qquad d = E(x|z=1) - c\ E(y|z=1) = \mu - c(\alpha + \beta\mu) = -c\alpha + (1-\pi)\mu,$$

so the implied estimator of α is

$$(26) \qquad a^* = -d/c = \alpha - \left(\frac{1-\pi}{\pi}\right) \beta\mu.$$

As at (14), the reverse regression estimator of α is downward biased. But now the bias persists even if the salary function is deterministic: $v = 0$ implies $\sigma_v^2 = 0$, but $\sigma_t^2 = \sigma_\varepsilon^2 + \sigma_v^2$ remains positive, so $\pi < 1$.

Nothing in the present specification implies that the absolute bias will be small. So $a^* < 0$ is quite possible even when $\alpha > 0$ (and $a = \alpha > 0$). Thus the reverse regression estimator may be negative even when discrimination is present. I suppose one could argue that a^* provides a lower bound for α, but so would any other statistic less than a, in this model.

Nothing in the present specification requires that ε in (17b) be interpreted as a random addition made by the employer to his productivity assessment. Rather it is intended to capture the additional productivity-relevant information that was available to the employer but not to the statistician. Our alternative causal model does require that the additional information be sex-free: $E(\varepsilon|z) = 0$.

Finally, what distinguishes (7) and (17) is not that (7) writes x as a function of p while (17) writes p as a function of x. Rather the distinction arises from the respective independence assumptions. In both models, p and z are correlated: $E(p|z) \neq 0$. In the errors-in-variable model (7) some of that correlation remains after controlling for measured qualifications. In the alternative causal model (17) none of it remains. As Bloom & Killingsworth (1982: 323) correctly point out, "The fact that an unmeasured variable that affects compensation is correlated with a measured variable...

does not necessarily mean...biased results. Rather, bias arises only if the unmeasured variable is correlated both with compensation and with a measured variable _at the margin_, i.e., when all other measured variables are held constant." Incidentally, this point is virtually identical to a critical issue in the earlier literature on estimating treatment effects when assignment to treatment and control groups is nonrandom: see Barnow & Cain (1977: 178-186).

Our alternative univariate model suffices to dispose of many claims and impressions. But a fuller evaluation of the virtues of reverse regression requires us to proceed to the multivariate case.

4. Multivariate Models

For the empirically relevant situation where there are several measured qualifications I develop three alternative models. In all three, the salary function will be deterministic: $y = p + \alpha z$. This simplification is made because it's most favorable for reverse regression; I will occasionally indicate informally results that hold up when the pure noise disturbance is restored to the salary function. In all three models, the qualification vector $\underline{x}$ will be imperfectly correlated with p, the latent variable which is best interpreted as the employer's assessment of productivity. The models differ with respect to their specification of the structural relationship between $\underline{x}$ and p. In Model A, which generalizes (17), the elements of $\underline{x}$ are causes of p. In Model B, which generalizes (7), the elements of $\underline{x}$ are indicators of p. Model C provides a distinct generalization of (7): the elements of $\underline{x}$ are, one for one, fallible measures of the elements of a true qualification vector $\underline{x}^*$, which in turn are causes of p. (In the univariate situation where $\underline{x}$ was scalar, the distinction between B and C did not arise). Figure 2 gives the path diagrams.

For each model, I derive, in terms of its structural parameters, the coefficients of the direct regression $E(y|\underline{x},z) = \underline{b}'\underline{x} + az$ and of the reverse regression system $E(\underline{x}|y,z) = \underline{c}y + \underline{d}z$. From these, the coefficients of the composite reverse regression $E(q|y,z) = cy + dz$ follow immediately: $q = \underline{b}'\underline{x}$ implies $c = \underline{b}'\underline{c}$ and $d = \underline{b}'\underline{d}$. The implied estimates of α, namely the $a_j^* = -d_j/c_j$ $(j=1, \ldots, k)$ and $a^* = -d/c$, follow as well.

In each model, the means, but not the variances and covariances, may differ by gender. Again I take all regressions to be linear-additive, so that the respective slope vectors can be calculated as

$$(27) \qquad \underline{b} = (V(\underline{x}|z))^{-1}C(\underline{x},y|z), \qquad \underline{c} = C(\underline{x},y|z)/V(y|z),$$

and then the gender coefficients follow as

$$(28) \qquad a = E(y|z=1) - \underline{b}'E(\underline{x}|z=1), \qquad \underline{d} = E(\underline{x}|z=1) - \underline{c}\,E(y|z=1).$$

It should be recognized that the main conclusions hold up under weaker assumptions provided that the regressions (conditional expectations) are reinterpreted as projections (best linear predictiors). Some of the derivations exploit two consequences of the salary function being deterministic: With $y = p + \alpha z$, for any variable T we have

$$(29a) \qquad E(y|T,z) = E(p|T,z) + \alpha z,$$

and

$$(29b) \qquad E(T|y,z) = E(T|p,z).$$

<u>Model A, "Multiple Causes"</u>. Suppose that

(30a) $\quad y = p + \alpha z$

(30b) $\quad p = \underline{\beta}'\underline{x} + w$

(30c) $\quad \underline{x} = \underline{\mu}\, z + \underline{u},$

with

(30d,e) $\quad E(\underline{u}|z) = \underline{0}, \qquad V(\underline{u}|z) = \Sigma$

(30f,g) $\quad E(w|\underline{x}, z) = 0, \qquad V(w|\underline{x}, z) = \sigma_w^2.$

Here the employer's assessment of productivity is determined by measured qualifications, subject to a gender-free disturbance. That disturbance represents the additional information available to the employer but not to the statistician. The means of $\underline{x}$ may differ by gender, a property which carries over to p and y. For interpretive purposes we may suppose that the elements of $\underline{\beta}$ and $\underline{\mu}$ are all positive: men rank higher than women on all the measured contributors to productivity.

From (30a, b, f) it follows immediately that

(31) $\quad E(y|\underline{x}, z) = \underline{\beta}'\underline{x} + \alpha z.$

Thus direct regression gives an unbiased assessment of discrimination ($a=\alpha$) despite the fact that the measured variables do not exhaust the information used by the employer in assessing productivity. (Appending a pure noise disturbance to (30a) would not affect that result).

To evaluate the reverse regression (and incidentally to verify the unbiasedness of direct regression), we calculate the within-gender moments of $\underline{x}$ and y:

(32a,b) $E(\underline{x}|z) = \underline{\mu}\, z, \qquad V(\underline{x}|z) = \Sigma,$

(32c,d) $E(y|z) = (\underline{\beta}'\underline{\mu} + \alpha)z, \qquad V(y|z) = \underline{\beta}'\Sigma\underline{\beta} + \sigma_w^2\,,$

(32e) $C(\underline{x}, y|z) = \Sigma\underline{\beta}.$

For the direct regression we verify $\underline{b} = \underline{\beta}$, $a = \alpha$. For the reverse regression system, on the other hand, we find

(33) $\underline{c} = (\underline{\beta}'\Sigma\underline{\beta} + \sigma_w^2)^{-1}\Sigma\underline{\beta}\,,$

(34) $\underline{d} = \underline{\mu} - \underline{c}\,(\underline{\beta}'\underline{\mu} + \alpha) = -\underline{c}\,\alpha + (I - \underline{c}\,\underline{\beta}')\,\underline{\mu}.$

So for the composite reverse regression,

(35) $c = \pi, \qquad d = -\pi\alpha + (1-\pi)\,\underline{\beta}'\underline{\mu},$

where

(36) $\pi = \underline{\beta}'\Sigma\,\underline{\beta}/(\underline{\beta}'\Sigma\,\underline{\beta} + \sigma_w^2)$

lies in the unit interval. The implied composite estimate of α is

(37) $a^* = \alpha - \left(\frac{1-\pi}{\pi}\right)\underline{\beta}'\underline{\mu}.$

Since $0 < \pi < 1$, a^* is downward biased. (The biases in the individual $a_j^* = -d_j/c_j$ are not determinate).

For a simple illustration of the bias of reverse regression, take k=2, and set

(38) $\alpha = 1,\ \underline{\beta} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \underline{\mu} = \begin{pmatrix} 3 \\ 2 \end{pmatrix},\quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\quad \sigma_w^2 = 1.$

Then

$$E(\underline{x}|z=1) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \qquad V(\underline{x}|z) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$E(y|z=1) = 6, \qquad V(y|z) = 3,$$

$$C(\underline{x},y|z) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

from which

(39) $\underline{c} = (1/3) \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \underline{d} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$

Here $a_1^* = -3$, $a_2^* = 0$. Each reverse regression points away from discrimination, as does the composite: $c = 2/3$, $d = 1 \Rightarrow a^* = -3/2$.

Model B, "Multiple Indicators". Suppose that

(40a) $y = p + \alpha z$

(40b) $\underline{x} = \underline{\gamma}\, p + \underline{\varepsilon}$

(40c) $p = \mu z + u,$

with

(40d,e) $E(u|z) = 0, \qquad V(u|z) = \sigma_u^2\ ,$

$$(40f,g)\quad E(\underline{\varepsilon}|p,z) = \underline{0}, \qquad V(\underline{\varepsilon}|p,z) = \Omega .$$

Here each observed qualification is merely an indicator of the employer's productivity assessment, subject to a gender-free disturbance. The mean p may differ by gender, a property that carries over to $\underline{x}$ and y. For interpretation we may suppose all elements of $\underline{\gamma}$ and μ are positive.

This causal model is my formulation of the instruction given by Roberts (1980: 180-181):

> "One ordinarily must be content with proxies or surrogate measures of productivity, which are often called job qualifications. Examples of qualifications are years of schooling, prior experience, seniority, and particular job skills.... Although ... a qualification could be something as concrete as years of education, it is best to think of it abstractly as simply a proxy for productivity."

At the same time it captures the "one-mediator null hypothesis" which supposes "that one factor, quality, underlies all the intercorrelations" among the observed variables: see Birnbaum (1979a, 1981, 1982). I am, however, not imposing the uncorrelated error requirement of classical factor analysis: the Ω matrix need not be diagonal.

The within-gender moments are

$$(41a,b)\quad E(\underline{x}|z) = \underline{\gamma}\,\mu\, z, \qquad V(\underline{x}|z) = \underline{\gamma}\,\underline{\gamma}'\,\sigma_u^2 + \Omega,$$

$$(41c,d)\quad E(y|z) = (\mu+\alpha)z, \qquad V(y|z) = \sigma_u^2 ,$$

$$(41e)\quad C(\underline{x},y|z) = \underline{\gamma}\,\sigma_u^2.$$

For the direct regression, the slope vector turns out to be

$$\underline{b} = \pi^*(\underline{\gamma}'\Omega^{-1}\gamma)^{-1}\Omega^{-1}\underline{\gamma}\,, \tag{42}$$

where

$$\pi^* = \sigma_u^2\, \underline{\gamma}'\Omega^{-1}\underline{\gamma}/(1 + \sigma_u^2\underline{\gamma}'\Omega^{-1}\underline{\gamma}) \tag{43}$$

lies in the unit interval, and then the gender coefficient is

$$a = \alpha + (1-\pi^*)\mu. \tag{44}$$

The direct estimator of α is biased upwards.

For the reverse regressions, on the other hand,

$$\underline{c} = \sigma_u^{-2}\, \underline{\gamma}\, \sigma_u^2 = \underline{\gamma}, \qquad \underline{d} = \underline{\gamma}\, \mu - \underline{c}(\mu+\alpha) = -\, \underline{\gamma}\, \alpha, \tag{45}$$

$$c = \pi^*, \quad d = -\pi^*\alpha, \tag{46}$$

whence $a_j^* = a^* = \alpha$. Thus all reverse regressions provide unbiased assessments of discrimination in the present model. This conclusion indeed follows directly from (40a,b,f):

$$E(\underline{x}|y,z) = E(\underline{x}|p,z) = \underline{\gamma}\, p = \underline{\gamma}(y-\alpha z) = \underline{\gamma}\, y - \underline{\gamma}\alpha z, \tag{47}$$

showing that $\underline{c} = \underline{\gamma}$, $\underline{d} = -\underline{\gamma}\alpha$, etc.

Our multiple-indicator model clearly supports Conway & Roberts's composite reverse regression $E(q\,|y,z) = cy + dz$ as a device for assessing discrimination. Indeed it justifies the use of any one of the separate reverse regressions, $E(x_j|y,z) = c_jy + d_jz$, for the same purpose. To put it in different words, the present model implies that in the reverse regression system, $E(\underline{x}|y,z) = \underline{c}y + \underline{d}z$, the vector $\underline{d}$ is proportional to the vector $\underline{c}$, the factor of proportionality being $-\alpha$.

In practice, where sampling variability is present, this proportionality restriction has obvious implications for model testing and efficient estimation, which I will examine in Section 5.

The pattern of restrictions which we have located in terms of regression coefficients was manifest in Birnbaum's (1979a: 124) display of implied correlations, although he did not dramatize it. Some readers will already have spotted the restrictions in terms of moments in our equations (41): The elements in $E(\underline{x}|z)$ -- which in view of our coding conventions represent the mean differences between the genders on the qualification variables -- are, in the present model, proportional to the corresponding covariances in $C(\underline{x},y|z)$.

For a simple illustration of the bias of direct regression, take

$$(48) \qquad \alpha = 1, \quad \underline{\gamma} = (1/6)\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \mu = 6, \quad \Omega = (1/6\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_u^2 = 6.$$

Then

$$E(\underline{x}|z = 1) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad V(\underline{x}|z) = (1/6)\begin{pmatrix} 10 & 6 \\ 6 & 5 \end{pmatrix},$$

$$E(y|z = 1) = 7, \qquad V(y|z) = 6,$$

$$C(\underline{x}, y|z) = \begin{pmatrix} 3 \\ 2 \end{pmatrix},$$

from which

$$(49) \qquad \underline{b} = (3/7)\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \qquad a = 10/7.$$

The direct regression overstates discrimination.

What happens if a disturbance v is added to the salary function (40a)? Taking v to have zero expectation and variance σ_v^2, independent of z, u, $\underline{\varepsilon}$, the only change in the moments is at (41d) which becomes $V(y|z) = \sigma_u^2 + \sigma_v^2 = \sigma_t^2$, say. Consequently the expressions for $\underline{b}$ and a are unaffected, while the reverse regression coefficients become

$$\underline{c} = \underline{\gamma}\,\phi\,, \qquad \underline{d} = \underline{\gamma}(-\phi\alpha+(1-\phi)\mu)$$

where $\phi = \sigma_u^2/\sigma_t^2$. Observe that the proportionality restriction persists. But none of the reverse regressions correctly assesses discrimination: $a_j^* = a^* = \alpha - (\frac{1-\phi}{\phi})\,\pi^*\mu$; all indeed are biased downward.

<u>Model C, "Errors in Variables".</u> Suppose that

$$(50a) \qquad y = p + \alpha z$$

$$(50b) \qquad p = \underline{\beta}'\underline{x}^*$$

$$(50c) \qquad \underline{x} = \underline{x}^* + \underline{\varepsilon}$$

$$(50d) \qquad \underline{x}^* = \underline{\mu}z + \underline{u}$$

with

$$(50e,f) \qquad E(\underline{u}|z) = \underline{0}, \qquad V(\underline{u}|z) = \Sigma^*,$$

$$(50g,h) \qquad E(\underline{\varepsilon}|\underline{x}^*,z) = \underline{0}, \qquad V(\underline{\varepsilon}|\underline{x}^*,z) =$$

Here the employer's productivity assessment is an exact function of a set of true qualification variables. Each observed qualification is a fallible indicator of the corresponding true qualification, the measurement errors being gender-free. The mean $\underline{x}^*$ may differ by gender, a property that carries over to p, $\underline{x}$, and y. For interpretation we may suppose all elements of $\underline{\beta}$ and $\underline{\mu}$ are positive.

This is my version of the model explicitly given by Hashimoto & Kochin (1980: 479-481) and Kamalich & Polachek (1982: 452-453). To make the situation more favorable for reverse regression, I've suppressed the disturbance in (50b). I have not imposed their uncorrelated error requirement (diagonal); doing so would not affect the form of my results.

The within-gender moments are

(51a,b) $E(\underline{x}|z) = \underline{\mu}\, z,$ $\qquad V(\underline{x}|z) = \Sigma^* + \quad = \Sigma$, say,

(51c,d) $E(y|z) = (\underline{\beta}'\underline{\mu}+\alpha)z,$ $\qquad V(y|z) = \underline{\beta}'\Sigma^*\underline{\beta}$

(51e) $C(\underline{x},y|z) = \Sigma^*\underline{\beta}.$

For the direct regression we deduce

(52) $\underline{b} = \Sigma^{-1}\Sigma^*\underline{\beta}$,

(53) $a = \alpha + (\underline{\beta} - \underline{b})'\underline{\mu},$

so the direct estimator is biased. I am unable to sign the bias in general, but as Hashimoto & Kochan (1980: 481) indicate, if is diagonal, then $a > \alpha$.

For the reverse regression system, we deduce

(54) $\underline{c} = (\underline{\beta}'\Sigma^*\underline{\beta})^{-1}\Sigma^*\underline{\beta}$,

$$(55) \qquad \underline{d} = -\underline{c}\alpha + (I - \underline{c}\ \underline{\beta}')\underline{\mu},$$

whence for the composite reverse regression,

$$(56) \qquad c = \pi, \qquad d = -\pi\alpha + (\underline{b} - \pi\underline{\beta})'\underline{\mu},$$

with

$$(57) \qquad \pi = \underline{b}'\underline{c} = \underline{\beta}'\Sigma^*\Sigma^{-1}\Sigma^*\underline{\beta}/\underline{\beta}'\Sigma^*\underline{\beta}$$

lying in the unit interval. Evidently $a_j^* \neq \alpha$: none of the reverse regressions provides an unbiased assessment of discrimination. Nor does the composite, for which

$$(58) \qquad a^* = \alpha - (\frac{1}{\pi}\underline{b} - \underline{\beta})'\underline{\mu}.$$

The directions the biases are indeterminate, even with Θ diagonal. Contrary to the claim of Kamalich & Polachek (1982: 456) the reverse regression system is <u>not</u> appropriate for estimating α in this multivariate errors-in-variable model.

For a numerical example, we take

$$(59) \qquad \alpha = 1, \quad \underline{\beta} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma^* = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$E(\underline{x}|z = 1) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad V(\underline{x}|z) = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix},$$

$$E(y|z = 1) = 4, \qquad V(y|z) = 3,$$

$$C(\underline{x}, y|z) = \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

from which

$$\underline{b} = (1/6) \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \qquad a = 7/3 ,$$

$$\underline{c} = (1/3) \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \underline{d} = (1/3) \begin{pmatrix} -5 \\ 2 \end{pmatrix} .$$

It follows that $a_1^* = 5/2$, $a_2^* = -2$, $a^* = 14/11$. Observe that, contrary to the bounding suggestion of Abowd et al. (1983), a_1^* and a^* are, along with a, larger than α in this example.

Postscript. Kamalich & Polachek (1982), whose model is explicitly of the multivariate errors-in-variable type, also propose and utilize a more elaborate version of reverse regression. As pointed out in subsection 3.2 above, in each reverse regression they include all other x's along with y and z as explanatory variables. That is, for j=1,...,k, they fit

(60) $$E(x_j|y, z, \underline{x}_j^o) = f_j y + g_j z + \underline{h}_j' \underline{x}_j^o$$

where $\underline{x}_j^o$ is $\underline{x}$ after deletion of x_j.

I take the implied estimators of α to be the $\hat{\alpha}_j = -g_j/f_j$. For each of our three models, one can deduce the coefficients in (60) in terms of its structural parameters. Having done so, I can report that the $\hat{\alpha}_j$ are all biased in Model C. Thus contrary to Kamalich & Polachek's claim (1982: 456), their more elaborate version of reverse regression is not appropriate for assessing discrimination in a multivariate errors-in-variable model. It is possible that sharper conclusions -- e.g. on the direction of bias -- can be obtained by use of the tools developed by Klepper & Leamer (1982), but I have not attempted that task.

Further, the estimates produced by (60) are biased under Model A, while under Model B they have the same qualitative properties as the simpler reverse regressions: namely, they are unbiased iff the salary function is deterministic, and they satisfy proportionality restrictions whether or not the salary function is deterministic.

Rather than develop these analytical conclusions, I will use the numerical examples to illustrate the results of applying (60).

For the Model A example (38), the regressions are

$$E(x_1|y,z,x_2) = (1/2)(y + 2z - x_2),$$
$$E(x_2|y,z,x_1) = (1/3)(y + z - x_2),$$

whence $\hat{\alpha}_1 = -2$, $\hat{\alpha}_2 = -1$. For the Model B example (48), the regressions are

$$E(x_1|y,z,x_2) = (1/2)\ (y-z)$$
$$E(x_2|y,z,x_1) = (1/3)\ (y-z)$$

whence $\hat{\alpha}_1 = \hat{\alpha}_2 = 1$. For the Model C example (59), the regressions are

$$E(x_1|y,z,x_2) = (1/5)\ (4y - 7z - x_2),$$
$$E(x_2|y,z,x_1) = (1/5)\ (y - 2x_1),$$

whence $\hat{\alpha}_1 = 7/4$, $\hat{\alpha}_2 = 0$. Recall that the true value is $\alpha = 1$ in all three examples, and that Kamalich & Polachek claimed that their procedure was justified for a specification of the type of Model C.

5. Model Discrimination.

The models developed above hardly exhaust the possibilities. It's easy enough to specify a general omitted-variable model in which the structural discrimination parameter is not identified so that neither direct nor reverse regression is appropriate. Nevertheless, we have reached a constructive conclusion. The only known stochastic specification under which reverse regression provides a consistent estimator of α is the multiple-indicator one, Model B. That model implies coefficient restrictions on the multivariate reverse regression system. In an empirical context, where sampling variability prevails, we can use the restrictions to test the validity of the model and thus the validity of the reverse regression estimators. And if the model is valid, we can use the restrictions to obtain a single optimal estimator of α.

I sketch the theory. For Model B, let

$$(61) \qquad \underline{s} = (y,z)', \quad \underline{\theta} = (1, -\alpha)', \quad \Pi = \underline{\gamma}\,\underline{\theta}',$$

so that

$$(62) \qquad E(\underline{x}|\underline{s}) = \Pi\underline{s}, \quad V(\underline{x}|\underline{s}) = \Omega.$$

This will be recognized as a classical multivariate regression model with rank-one restriction on the coefficient matrix. We suppose that the sample consists of independent observations. If we add the assumption that $\underline{x}|\underline{s}$ is multinormal, our model is precisely of the type considered by Anderson (1951), Hauser & Goldberger (1971), and Leamer (1978: 243-253). Leamer indeed explicitly discusses reverse regression (p. 252). Maximizing the likelihood function is accomplished by extracting a certain characteristic vector (which serves to estimate $\underline{\gamma}$ and α) and concurrently producing the characteristic root which enters the likelihood-ratio test statistic. In the absence of normality, the same parameter estimators retain desirable properties since they also follow from a minimum-distance principle.

Thus in practice one can draw on standard principles to discriminate Model B from its competitors. When Model B is statistically rejected, I see no scientific basis for using reverse regression to assess salary discrimination. It would be nice to carry out this program for all the data sets discussed previously. I haven't yet done so. But working from the published articles and from unpublished materials kindly provided by Professor Conway and Professor Polachek, I have reconstructed most of the relevant moments (sample means, variances, and covariances of the observed variables), and so can report some results.

For Conway & Roberts's (1983) Chicago bank sample, for which the direct and composite-reverse regressions were given in Table 3, I've calculated the six separate reverse regressions and six conflicting estimates of α that they imply. These are given in Table 5, along with the gender mean differences on the observed variables (sample analogues of $E(\cdot|z)$) as background and the composite reverse results as repetition. To the naked eye the wildly different a_j^*'s suggest that the rank-one restriction is invalid. But the likelihood-ratio test statistic is 6.8 which is not a surprising value from a $\chi^2(5)$ distribution, so Model B is acceptable. The ML estimate of α is .020, quite similar to the composite estimate $a^* = .031$.

For a number of reasons, the Conway & Roberts study is an awkward choice to illustrate the statistical assessment of salary discrimination: the coding of the x-variables is unclear, the set of x's includes a squared term and an interaction term but not the underlying linear terms, and the within-gender moments suggest that interactions with gender may be present (despite Conway & Roberts's (1983: 77) reassurances about diagnostic checking). Some readers may also be bemused by the notion that gender discrimination is being addressed with respect to salary for a group of 237 employees of whom only 37 are female.

For Kamalich & Polachek's (1982) national sample, for which the direct and separate reverse regressions were given in Table 2, my results should be taken as tentative because of the slippage involved in reconstructing moments from the rounded figures available to me.

The upper panel of Table 6 refers to gender. The gender mean differences are tabulated, and the separate reverse regression coefficients are repeated from Table 2. The new entries are the implied estimates $a_j^* = -d_j/c_j$, and the composite reverse regression results. The substantial differences among the a_j^* suggest that the rank restriction is invalid. If the restrictions are valid, then the ML estimate of α is .24, essentially the same as the composite estimate $a^* = .26$. But the likelihood-ratio test statistic is 172 which is a very surprising value on the null hypothesis that the restrictions are valid. (The null distribution is $\chi^2(2)$). By conventional standards of statistical inference, therefore, Kamalich & Polachek's reverse regressions are useless as assessments of salary discrimination against women in their sample.

The situation is similar for race, to which the lower panel of Table 6 refers. Again the a_j^* span a wide range. Subject to the restrictions, the

ML estimate of α is $-.39$, which is quite close to the composite estimate, namely $a^* = -.31$. But the likelihood-ratio test statistic is 122, a very surprising value from the relevant null distribution, namely $\chi^2(2)$. By conventional standards of statistical inference, therefore, Kamalich & Polachek's reverse regressions are useless as assessments of salary discrimination against (or,for that matter, in favor of) blacks in their sample.

6. Concluding Remarks

I conclude that reverse regression results should not be taken seriously unless accompanied by the information needed to test the restrictions of the muliple-indicator model. That model, to repeat, is the only one that can support the new approach for estimating the discrimination parameter as defined here. I have ignored a quite distinct rationale for the new approach, introduced by Conway & Roberts (1983). Their reading of ethical principles is that, regardless of stochastic specification, the coefficients of the reverse regressions are legitimate parameters of interest. For critical perspectives on the ethical, legal, and statistical issues, see Fisher (1980), Finkelstein (1980: 747-749), Michelson & Blattenberger (1983), Weisberg & Tomberlin (1983), Greene (1983), and several of the contributions to a forthcoming symposium in the Journal of Business & Economic Statistics.

To focus on certain statistical issues, I've relied on a very primitive view of salary determination. Consequently the important issues of compensation packages, information dynamics, hiring, promotion, and retention have been ignored here, as in most of the reverse regression literature. Some of these issues can be handled within a selectivity-bias framework, as has been done by Abowd & Killingsworth (1982), Abowd, et al. (1983), and Abowd (1983).

Figure 1. Path Diagrams for Two Models

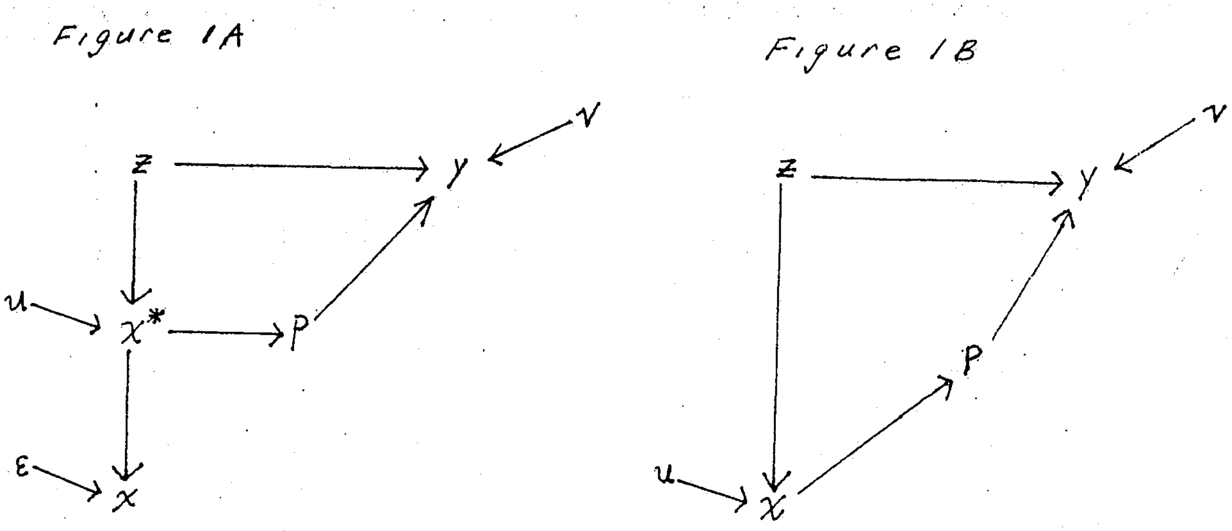


Table 1

Income and Education (1959)
Males (25+) Years of Age

(A)
Median Income by Schooling
Schooling (Years)

	None	1-4	5-7	8	High School 1-3	High School 4	College 1-3	College 4+
White	1569	1962	3240	3981	5013	5529	6104	7779
Non-White	1042	1565	2353	2900	3253	3735	4029	4840

(B)
Median Schooling by Income
Income ($1,000s)

	None	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-9	10+
White	8.4	8.0	8.4	8.7	9.5	10.5	11.4	12.1	12.4	14.0
Non-White	6.9	5.1	6.5	7.8	8.7	9.3	10.4	11.2	12.1	12.8

Source: U.S. Census of Population 1960 (Vol. 1, Part 1, Table 223). United States Census Bureau, 1960, as appears in Hashimoto & Kochin (1980).

TABLE 2. DIRECT & REVERSE REGRESSION COEFFICIENTS IN THE 1976 PANEL STUDY OF INCOME DYNAMICS

$y = \ell n$ wage, x_1 = education (years), x_2 = tenure (months), x_3 = experience (years).

Gender: $z = 1$ if male, $= 0$ if female.

Direct Regression		Reverse Regressions			
Dependent var:	y	Dep. var:	x_1	x_2	x_3
x_1	.075	y	2.87	51.06	3.41
x_2	.0012	z	−1.44	13.48	3.23
x_3	.0051				
z	.351				

Race: $z = 1$ if white, $= 0$ if black.

Direct Regression		Reverse Regressions			
Dependent var:	y	Dep. var:	x_1	x_2	x_3
x_1	.067	y	1.84	59.10	5.26
x_2	.0014	z	1.46	−8.06	−1.63
x_3	.0067				
z	.133				

Source: Adapted from Kamalich & Polachek (1982: 452, 458) and unpublished material provided by Professor Polachek.

TABLE 3: DIRECT & REVERSE REGRESSION COEFFICIENTS (AND STANDARD ERRORS) FOR A CHICAGO BANK

y = ℓn salary; x_1, x_2, x_3 = "categorical variables for educational levels"; x_4 = months of work experience prior to hire; x_5 = = square of seniority in months; x_6 = "an interaction variable created from "x_4 & age; z = 1 if male, 0 if female.

Direct Regression		Reverse Regression	
Dependent var: y		Dependent var: $q = \underline{b}'\underline{x}$.	
x_1	.1844 (.0838)	y	.316 (.0282)
x_2	.4427 (.0764)	z	−.0097 (.0202)
x_3	.5647 (.0782)		
x_4	−.0006 (.0002)		
x_5	.0109 (.0020)		
x_6	−3.4917 (.7309)		
z	.1482 (.0356)		
R^2 = .378		R^2 = .329	

Source: Adapted from Conway & Roberts (1983: Tables 2 & 3), and correction supplied by Professor Conway.

TABLE 4. IMPLIED COEFFICIENTS (AND STANDARD ERRORS) ON GROUP STATUS FROM DIRECT & REVERSE REGRESSIONS IN 1976 SURVEY OF INCOME & EDUCATION

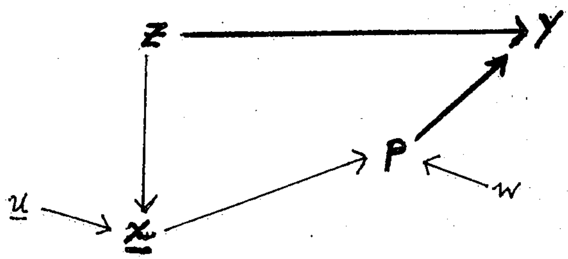
y = ℓn wage; x = qualification vector; z = 1 if white, 0 if indicated minority group.

	Federal Sector			Non-Federal Sector		
	Puerto Rican	Other Hispanic	Black	Puerto Rican	Other Hispanic	Black
Direct:	.1279 (.0821)	.0476 (.0337)	.1612 (.0204)	.0697 (.0482)	.0080 (.0241)	.1420 (.0202)
Reverse:	.0056 (.0521)	.0945 (.0195)	.0390 (.0116)	.0818 (.0211)	.0682 (.0090)	.0093 (.0095)

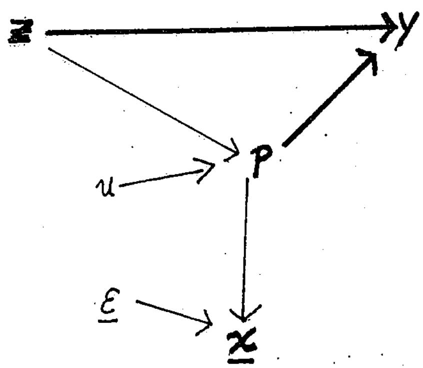
Source: Adapted from Abowd, Abowd, & Killingsworth (1983: Table 3, panel (B)i). Dependent variable: y in direct regression, q = b'x in reverse regression. Coefficients are actually differentials obtained from separate regressions for whites & minority groups, evaluated at white means.

Figure 2. Path Diagrams for Three Multivariate Models

MODEL A
Multiple Causes



MODEL B
Multiple Indicators



MODEL C
Errors in Variables

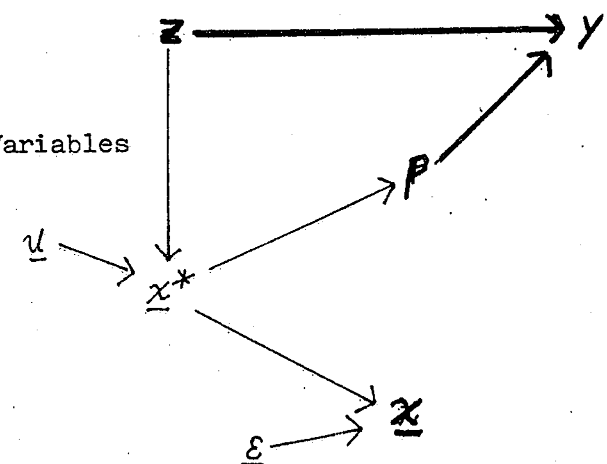


TABLE 5. ANALYSIS OF REVERSE REGRESSIONS FOR A CHICAGO BANK

(See Table 3 for definitions of variables).

j	Variable	Mean Difference	Reverse Regression Estimates c_j	d_j	a_j^*
1	x_1	.0033	-.35	.074	.21
2	x_2	-.124	-.098	-.104	-1.06
3	x_3	.181	.595	.060	-.10
4	x_4	9.07	-56.94	20.60	.36
5	x_5	3.88	3.06	3.26	-1.07
6	x_6	.0088	-.0066	.0101	1.54
	z	1.00	--	--	--
	y	.203	--	--	--
	q	.054	.316	-.0097	.031

Source: See Table 3 and text.

TABLE 6. ANALYSES OF REVERSE REGRESSIONS IN THE 1976 PANEL STUDY OF INCOME DYNAMICS

(See Table 2 for definitions of variables).

GENDER

j	Variable	Mean Difference	Reverse Regression Estimates c_j	d_j	a_j^*
1	x_1	-.33	2.87	-1.44	.50
2	x_2	33.33	51.06	13.48	-.26
3	x_3	4.56	3.41	3.23	-.95
	z	1.00	--	--	--
	y	.39	--	--	--
	q	.04	.295	-.075	.26

RACE

j	Variable	Mean Difference	c_j	d_j	a_j^*
1	x_1	1.97	1.84	1.46	-.79
2	x_2	8.20	59.10	-8.06	.14
3	x_3	-.19	5.26	-1.63	.31
	z	1.00	--	--	--
	y	.27	--	--	--
	q	.14	.241	.076	-.31

Source: See Table 2 and text.

REFERENCES

A. Abowd (1983), Pay Comparisons between Federal and Non-Federal Employers and Between Blacks and Hispanics using Selectivity Adjustments and Reverse Regression, University of Chicago: Graduate School of Business, doctoral dissertation, June.

A. M. Abowd, J. M. Abowd, & M. R. Killingsworth (1983), "Race, Spanish origin and earnings differentials among men: the demise of two stylized facts", University of Chicago: Economics Research Center/NORC Discussion Paper 83-11, May, 46 pp.

J. M. Abowd & M. R. Killingsworth (1982), "Employment, wages and earnings of Hispanics in the Federal and non-Federal sectors: methodological issues and their empirical consequences," January, 66 pp.

T. W. Anderson (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions," Annals of Mathematical Statistics, 22, 327-351.

D. C. Baldus & J.W.L. Cole (1980), Statistical Proof of Discrimination, New York: McGraw-Hill. (+ 1982 Cumulative Supplement).

B. S. Barnow & G. G. Cain (1977), "A reanalysis of the effect of Head Start on cognitive development: methodology and empirical findings," Journal of Human Resources, 12, 177-197.

M. H. Birnbaum (1979a), "Procedures for the detection and correction of salary inequities," pp. 121-144 in T. R. Pezzullo & B. E. Brittingham, editors, Salary Equity, Lexington, MA: Lexington Books.

M. H. Birnbaum (1979b), "Is there sex bias in salaries of psychologists?", American Psychologist, August, 719-720.

M. H. Birnbaum (1981), "Reply to McLaughlin: proper path models for theoretical partialling," American Psychologist, October, 1193-1195.

M. H. Birnbaum (1982), "On the one-mediator null hypothesis of salary equity," American Psychologist, October, 1146-1147.

D. E. Bloom & M. R. Killingsworth (1982), "Pay discrimination research and litigation: the use of regression," Industrial Relations, 21, Fall, 318-339.

D. A. Conway & H. V. Roberts (1983), "Reverse regression, fairness, and employment discrimination", Journal of Business & Economic Statistics, 1, January, 75-85.

A. P. Dempster (1982), "Alternative models for inferring employment discrimination from statistical data," Harvard University: Department of Statistics Research Report, September, 34 pp.

M. O. Finkelstein (1980), "The judicial reception of multiple regression studies in race and sex discrimination cases," Columbia Law Review, 80, September, 737-754.

F. M. Fisher (1980), "Multiple regression in legal proceedings," Columbia Law Review, 80, September, 702-736.

M. Friedman (1953), A Theory of the Consumption Function, Princeton, NJ: Princeton University Press.

S. Garber & S. Klepper (1980), "On the diffusion of inconsistencies in linear regression," Economics Letters, 6, 125-129.

William F. Greene (1983), "Reverse regression, the algebra of discrimination," New York University: Graduate School of Business Administration, June, 13 pp.

M. Hashimoto & L. Kochin (1980), "A bias in the statistical estimation of the effects of discrimination," Economic Inquiry, 18, July, 478-486.

R. M. Hauser & A. S. Goldberger (1971), "The treatment of unobservable variables in path analysis," pp. 81-117 in H. L. Costner, editor, Sociological Methodology 1971, San Francisco: Jossey-Bass.

L. G. Humphreys (1981), "Theoretical partialling requires a defensible theory," American Psychologist, October, 1192-1193.

R. F. Kamalich & S. W. Polachek (1982), "Discrimination: fact or fiction? An examination using an alternative approach," Southern Economic Journal, 49, October, 450-461.

C. Kapsalis (1982), "A new measure of wage discrimination," Economics Letters, 9, 287-293.

S. Klepper & E. E. Leamer (1982), "Consistent sets of estimates for regressions with errors in all variables," UCLA: Department of Economics Discussion Paper 282, December, 37 pp.

E. E. Leamer (1978), Specification Searches, New York: John Wiley & Sons.

B. G. Malkiel & J. A. Malkiel (1973), "Male-female pay differentials in professional employment," American Economic Review, 63, September, 693-705.

G. P. McCabe, Jr. (1980), "The interpretation of regression analysis results in sex and race discrimination problems," American Statistician, 34, November, 212-215.

S. Michelson & G. Blattenberger (1983), "Reverse regression analysis of wage discrimination: a critique," Washington, DC: Econometric Research Inc., March 16, 36 pp.

R. Oaxaca (1973), "Male-female wage differentials in urban labor markets," International Economic Review, 14, October, 693-709.

H. V. Roberts (1980), "Statistical biases in the measurement of employment discrimination", pp. 173-195 in E. R. Livernash, editor, Comparable Worth: Issues and Alternatives, Washington, D.C.: Equal Employment Advisory Council.

H. Robbins & B. Levin (1983), "A note on the underadjustment phenomenon," Statistics and Probability Letters, 1, 137-139.

G. Solon (1983), "Errors in variables and reverse regression in the measurement of wage discrimination," Economics Letters, 13, 393-396.

H. I. Weisberg & T. J. Tomberlin (1983), "Statistical evidence in employment discrimination cases," Sociological Methods & Research, 11, 381-406.

L. Wolins (1978), "Sex differentials in salaries: faults in analysis of covariance," Science, 200, 19 May, 717.