©Journal of Business & Economic Statistics, Vol. 1, No. 1, January 1983

Reverse Regression, Fairness, and Employment Discrimination

Delores A. Conway and Harry V. Roberts
Graduate School of Business, University of Chicago, Chicago, IL 60637

Possible salary discrimination can be studied by comparing mean salaries of, say, males and females, after statistical adjustment for differences in job qualifications. The adjustment is often made by regression, with salary as dependent variable, and job qualifications and sex as independent variables. One might also regress job qualifications on salary and sex, a procedure called *reverse regression*. Ideas about fairness as well as technical concepts are relevant to discrimination studies. There are two distinct aspects of fairness, one based on comparisons of salary and the other based on comparisons of qualifications. Both concepts are needed to evaluate fairness.

KEY WORDS: Regression phenomenon; Fairness 1; Fairness 2; Affirmative action; Shunting.

1. INTRODUCTION

In studies of possible employment discrimination in an organization, we often have data for a group of employees that can be thought of as follows:

1. Y is a measure of pecuniary or nonpecuniary income such as a salary, job offer, or a promotion. Thorough studies of discrimination should consider all important dimensions of income, but actual litigation has typically been more narrowly focused, usually on pecuniary income of those actually hired, or rates of hiring from applicant pools. (Questions of initial selection of candidates for jobs can be brought within the income framework by regarding initial consideration for a job as a form of income to potential job holders.)
2. X is a composite index of "recorded job qualifications." In our exposition, X is often assumed univariate for simplicity of discussion. There may be errors of measurement in the recorded job qualifications that define X, and there are other potentially relevant qualifications that are not available. Hence, X is an imperfect measure of actual productivity, but we shall mainly think of X as a job qualification rather than as a proxy for productivity.
3. *Sex* is an indicator variable that takes the value 1 for females and 0 for males, or 1 for minorities and 0 for whites. For simplicity of exposition, we shall consider study of possible sex discrimination rather than racial or other forms of discrimination.

When concern focuses on possible salary discrimination, statisticians have tended to study the conditional distribution of Y given X and *Sex*, that is, to regress Y on X and *Sex*. The regression coefficient of *Sex* estimates mean salary differences between females and males after statistical allowance for the measured qualifications X. This regression corresponds to the idea that discrimination has to do with disparity in mean salaries for given measured job qualifications.

Another type of possible discrimination is placement discrimination, which refers to the "shunting" or "steering" of females or minorities into lower job levels than their qualifications warrant. In this case, Y is a categorical variable that represents the entering job group, where the job groups are numbered ordinally from lower to higher levels of skill, knowledge, responsibility, challenge, and perquisites. In this application, it is more natural to compare the average qualifications of males and females within each job group, that is, to regress X on Y and *Sex*. This regression corresponds to the idea that discrimination has to do with disparities in mean measured qualifications for given entering job groups.

A natural starting point for studies of employment discrimination is to compare the income and qualifications of males with those of females to determine whether or not one group has been treated unfairly. Here we refer to unfairness of the observed result seen in the data, not to a process assumed to have given rise to the observed data. We begin by focusing on the question: "What comparisons are appropriate for assessing fairness?"

To establish a basis for comparison, it is helpful to condition on either income or qualifications. We refer to comparisons based on the conditional distribution of Y given X as direct regression methods and those based on the conditional distribution of X given Y as reverse regression methods. Both methods are relevant for the study of possible discrimination of any kind. For example, in studies of salary discrimination, one can compare salaries at given measured qualifications (direct regression) or qualifications at given salaries (reverse regression). Similarly, in studies of placement discrimination, one can compare the probabilities of placement in particular job groups at given measured qualifications or the average measured qualifications within each job group. Both comparisons are appropriate, even though, as in the previous illustrations of salary and placement discrimination, one or the other may initially appear more natural.

Direct and reverse regression comparisons provide two perspectives of fairness:

1. The distributions of male and female incomes are the same at given qualifications. (This corresponds to direct regression.)
2. The distributions of male and female qualifications are the same at given incomes. (This corresponds to reverse regression.)

People unfamiliar with the statistical idea of regression often have trouble seeing that these two perspectives are different. When one views employment data bearing on possible discrimination from the two perspectives, the conclusions about fairness can be substantially different. In Section 2, we present actual data from an employment study that illustrates the different conclusions about fairness suggested by each perspective.

The choice between the two perspectives, or the consideration of both, points to problems of defining discrimination that have largely been ignored or misunderstood. The issues turn in part on questions that are well studied by simple numerical examples, which lawyers call "hypotheticals." In Sections 4 and 5 of this article, we display a series of hypotheticals that bring out problems of definition of discrimination. To make the ideas concrete, we consider discrimination against females. However, the same concepts apply to other protected classes, such as minority groups. Our intent in these sections is not to argue in favor of one viewpoint or the other, but to demonstrate the subtlety of the choice. We believe that questions about fairness have been at the root of much of the disagreement about the interpretation of data in discrimination studies.

The main conclusion of the article is that examination of the data in employment studies of discrimination from both perspectives—direct and reverse regression—is necessary to help decide whether males and females have been treated fairly. Both direct and reverse regression provide appropriate comparisons for assessing fairness.

In focusing on ideas of fairness of observed results, we deliberately omit or consider only in passing other important econometric and statistical questions that occur in studies of employment discrimination. One such question involves statistical significance of results. Other practical problems that must be addressed in discrimination studies include such issues as definition of relevant job qualifications and income variables, data cleaning, treatment of outliers, study of interaction effects between sex and qualifications, specification of statistical models, assessment of model fit and adequacy, and diagnostic checking of model results. An extended discussion of many of these items can be found in Roberts (1979).

Other important issues pertain to the development of models that specify the structure or causal mechanisms presumed to underly the observed data. In actual applications, models for inferring the presence of discrimination range from simple descriptive statistics to complete econometric formulations of different aspects of the employment process. The specification of such models involves economic and statistical issues that go beyond the concepts of fairness developed in this article. Economic concepts help to formulate structural models of the employment process whereas statistical concepts help to determine whether the data are in agreement with specific models. Although these issues are important for the implementation of direct and reverse regression comparisons, they are only briefly considered in Section 6. We hope to discuss more fully in a forthcoming paper some of the ways in which questions of underlying structure affect direct and reverse regression comparisons.

Since different perspectives of fairness largely precede questions of underlying economic structure and model specification, and since fairness has received little explicit attention previously, our main focus is on fairness. It is of course possible that the concept of fairness can conflict in many ways with economic efficiency, so that fairness cannot be completely separated from other goals. But in affirmative action, as in other areas of controversy about public policy, ideas of fairness—often called "equity" and contrasted with "efficiency"—may dominate other considerations.

2. AN APPLICATION

Readers may be helped by a simple application of direct and reverse regression comparisons that illustrates the different perspectives of fairness provided by each viewpoint. We give brief excerpts from an employment study of 274 professional employees hired at a large bank between 1971 and 1972. (This application is an excerpt from a much larger ongoing study, and is intended only to suggest how direct and reverse regression may be used in concrete applications.)

Professional employees were hired into seven job grades, ranging from 3 to 9. A simple application of reverse regression is to compare the mean educational levels between males and females across the seven entering job groups. Table 1 presents these comparisons,

Table 1. Number of Employees and Average Educational Level by Entering Job Group and Sex for Professional Employees Hired at a Large Bank during 1971–1972

Entering Job	Number of Employees		Average Education	
	Male	Female	Male	Female
3 Analytical	50	14	7.6	7.7
4 Technical	0	0	—	—
5 Systems	49	6	8.1	7.5
6 Administrative	24	14	7.6	7.6
7 Sales	20	0	8.3	—
8 Lending	61	0	8.4	—
9 Management	33	3	7.8	6.7
Total	237	37	8.0	7.6

based on a scaled variable for educational level. (On this scale, 4 refers to a high school degree, and 8 refers to a bachelor's degree.) The data in this table constitute a simple shunting study done from the perspective of reverse regression.

For entering job group 3, there are 50 males and 14 females. The females have slightly higher mean education, 7.7, as compared with 7.6 for the males. In the remaining three classes where both males and females are represented, the males have higher mean education in groups 5 and 9, while group 6 is a standoff. Overall, the 237 males have mean education 8.0, while the 37 females have mean education 7.6.

Next we show an example of direct regression in salary comparisons for males and females after statistical adjustment for job qualifications. For the same group of employees considered in Table 1, the natural log of 1976 salary (*LSAL*76) was regressed on four available job classifications (which we termed "basic qualifications"), linearizing transformations thereof, and *SEX*. Summary measures for the preferred model appear in Table 2. (In the larger study from which this example was taken, we found that details of model specification had small effects on the outcomes. The general approach to model search for a similar set of data is described in Roberts (1979). This report gives a much wider range of findings for another bank, and technical details on statistical methodology, including model selection, diagnostic checking, and policy on statistical significance.)

Here *ED*7, *ED*8, and *ED*9 are categorical variables for educational levels; WORK is number of months of prior work experience prior to hire; SENSQ is the square of seniority in months; *WK/AGE* is an interaction variable created from *WORK* and *AGE*.

The sex coefficient is −.148 with a standard error of .0356, suggesting an estimated salary shortfall of 13.8 percent on direct regression (recall that salary is expressed in natural log units). As in many applications of direct regression, nominal statistical significance is clear.

To implement the reverse regression corresponding to this direct regression, we create a variable *FITADJ* by using fitted values from direct regression but omitting

Table 2. Direct Regression Results of Salary on Job Qualifications for the Professional Employees in Table 1

Variable	B	Std. Error(B)	T
SEX	−.1482	.0356	−4.162
ED7	.1844	.0838	2.201
ED8	.4427	.0764	5.792
ED9	.5647	.0782	7.221
SENSQ	−.0006	.0002	−3.425
WORK	.0109	.0020	5.361
WK/AGE	−3.4917	.7309	−4.777
CONSTANT	12.3650	.8412	14.700

	Multiple R	R-Squared
Unadjusted	.609	.371
Adjusted	.595	.354

Standard Deviation of Residuals = .1968
N = 274

Table 3. Reverse Regression Results of Qualification Index on Salary for the Professional Employees in Table 1

Variable	B	Std. Error(B)	T
SEX	.0097	.0202	.479
LSAL76	.0316	.0282	11.192
CONSTANT	6.8018	.2810	24.202

	Multiple R	R-Squared
Unadjusted	.574	.329
Adjusted	.570	.324

Standard Deviation of Residuals = .1097
N = 274

the contribution of the *SEX* variable. *FITADJ* can be thought of as an index of qualifications, with weights determined by the relation between log salary and qualifications within each of the two sex groups. Then *FITADJ* is regressed on *SEX* and *LSAL*76. The corresponding summary measures are given in Table 3. (When there is only one qualification, the reverse regression meets the standard assumptions of linearity and constant dispersion. In applications like this one, diagnostic checking suggests that reverse regressions tend to be about as well specified as the direct regressions from which they were derived. The major problem we have encountered is due to the discreteness of the years-of-schooling variable, but the effect on conclusions from reverse regression appears to be small.) The point estimate for *SEX* is .01 with a standard error of .02. The females have a slight, but nonsignificant, excess of qualifications.

Hence, 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. The models appear to be well specified, and we have encountered similar findings in several sets of data. Moreover, the finding of approximate qualifications parity on reverse regression, coupled with female (or minority) salary shortfalls on direct regression has been established by other studies (see, e.g. Birnbaum, 1979a,b).

3. STATISTICAL DEFINITION OF DISCRIMINATION

Loosely speaking, employment discrimination refers to an unfair distribution of pecuniary or nonpecuniary income between, say, females and males in an organization. An example of pecuniary income is a starting salary, whereas the nature of a starting job is an example of nonpecuniary income. Consideration of both forms of income is desirable in any comprehensive study of possible discrimination.

As pointed out earlier, fairness has to do with the relation of income and qualifications. Ideas as to what is fair and unfair are unavoidable in dealing with the subject of discrimination. What has not been obvious is that there is more than one way to view the relationship between income and qualifications in order to address the concept of fairness.

3.1 Discrimination Defined Operationally by Direct Regression

Formal definitions of discrimination are hard to find in legal writings, statutes, administrative regulations, and judicial decisions. There appears to be an implicit assumption that discrimination is easier to recognize than to define. Schlei and Grossman (1976, Ch. 36) point out that the main focus of employment discrimination litigation has been on the consequences of discriminatory employment practices. They further discuss the use of statistical evidence as proof of inequitable treatment of a race or sex group, but do not formalize the concept with which statistical evidence is supposed to deal.

When the point of contention concerns discrimination in wages, salaries of male and female employees of a firm are examined for proof of disparate impact. Sometimes, discrimination is naively identified with a simple difference between male and female mean salaries. This viewpoint ignores the fact that salary differences are directly related to differences in job qualifications among workers, that should not be confounded with discrimination. In fact, Title VII of the Civil Rights Act of 1964 specifically mentions the use of seniority, employee merit, quantity of work, and quality of production as valid reasons for disparate treatment (Gwartney et al. 1979, p. 639). Current statistical studies of employment discrimination attempt to adjust salaries so as to separate the effects due to differential job qualifications from those that are due to employer discrimination. Adjustments are often made by direct regression methods, which means ordinary least squares regression or any other technique, such as cross-tabulation, designed to study the conditional distribution of income given job qualifications.

Finkelstein (1980) and Fisher (1980) discuss the use of linear regression models in legal cases of employment discrimination. A typical model defines the dependent variable Y as a measure of income, such as salary or log salary, and the independent variables X as measures of job qualifications, such as years of schooling, work experience, or special skills related to the job. A common direct regression approach is to regress Y on the X's and an indicator variable, *Sex*, that assumes the value 1 for females and 0 otherwise. The direct regression model is given by:

$$E(Y \mid X) = a_1 + b_1'X + c_1 Sex \qquad (1)$$

It may be necessary to transform Y or the X's to achieve closer conformity to the specifications of the regression model, such as the linearity assumption. The model specification implies that the b vector is the same for males and females, that is, there is no sex by qualification interaction. (In our experience, interaction effects have been slight. If they are not slight, the problems of interpretation become severe not only for statisticians but for judges. Pronounced interaction effects could lead to data patterns in which any concept of simple discrimination or unfairness against females of minorities is lost. Some females or minorities may appear to benefit from discrimination while others may appear to suffer from it.) If diagnostic checking supports model adequacy, it is assumed that evidence about discriminatioin is conveyed by the estimated coefficient of the sex variable and its standard error. The sex coefficient represents the estimated difference between female and male average salaries, after statistical adjustment for the effects of the job qualifications, X. (In actual employment studies, many practical complications arise that must be addressed. For example, does the salary measure Y take into account hours of work, overtime pay, or bonuses?)

When this model is applied to employment data and some allowance is made for statistical significance, a common conclusion is that males on average receive higher salaries for given job qualifications than do females. If no measured job qualifications other than X are available, the conclusion reached is that the data show evidence of discrimination against females. A suggestive legal expression is "prima facie case of discrimination." This expression means that the defendant may wish to rebut the prima facie evidence by some appeal to additional information, presumably job qualifications not yet taken into account. At least until rebuttal is attempted, the prima facie case is often regarded as an operational definition of discrimination.

3.2 Limitations of the Operational Definition

The direct regression approach has been widely used in studies of employment discrimination. In fact, the statistical definition of discrimination proposed by Weisberg and Tomberlin (1982) is equivalent to direct regression. But there are difficulties with this operational definition of discrimination.

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. Rarely is any measure of performance included among available qualifications. (Performance ratings of employees by supervisors are, with reason, regarded as "tainted" evidence and are usually excluded from statistical studies in litigation about discrimination.) One may therefore ask whether there are other legitimate job qualifications, not captured in the available X variables, that ought to be used in statistical adjustment.

The existence of missing job qualifications raises the possibility that the statistical verdict at any stage of a study can be reversed by further evidence. This possibility is easy to forget or to dismiss. The problem of omitted job qualifications points to the weakness of a direct-regression-adjusted income differential as a definition of discrimination. The definition of discrimination depends on which job qualifications happen to be measured in a particular application.

But in evaluating the evidence available at any time, the job qualifications in the data base are all we have. Judgments about the effects of omitted variables must draw on information other than that contained in the data base of the study. These judgments can range from simple common-sense conjectures to econometric models of underlying structure or causal mechanisms. (Econometric models that pertain to errors-in-the-variables are clearly relevant, since they can be used to analyze the direction of bias due to left-out variables.)

For every direct regression model, there is a corresponding reverse regression model that brings out another aspect of the data. As suggested in Section 2, the perspective of direct regression often suggests the appearance of discrimination while the perspective of reverse regression suggests the appearance of nondiscrimination. A definition of discrimination that is based on direct regression alone fails to recognize and reconcile the alternative perspective provided by reverse regression.

3.3 Reverse Regression: An Alternative Perspective

Reverse regression is directed to the conditional distribution of measured job qualifications given income, rather than of income given measured job qualifications. The name suggests that the usual roles of Y and X are reversed. For a single job qualification variable, X, the reverse regression model is given by

$$E(X \mid Y) = a_2 + b_2 + c_2 Sex, \tag{2}$$

where *Sex* is an indicator variable that assumes the value 1 for females and 0 for males. For the case of multiple job qualification variables, we first construct a univariate qualification index $Q = a_1 + b_1{}'X$ using the estimated coefficients from the direct regression model given by (1). The qualification index differs from the fitted values of the direct regression in that the female indicator variable is not included. The reverse regression model is then given by (2) with Q substituted for X. This procedure amounts to a simplification of what is in principle a multivariate regression, since the dependent variable, "job qualifications," is multivariate.

Because of the conditioning on salary, reverse regression exposes features of the data that might go undetected by the use of direct regression alone. Comparison of employee qualifications for given jobs or salaries is especially pertinent when one is trying to detect shunting, which is usually thought of as placement of members of protected classes into lower job groups than would be suggested by their qualifications. In Section 4, we show how the perspectives of direct and reverse regression can suggest different conclusions about discrimination. The juxtaposition of direct and reverse regression makes one realize that the concept of fairness, usually taken to be self-evident, has two dimensions rather than one. The two dimensions do not necessarily suggest the same practical conclusion.

Hence direct and reverse regression, used in tandem, help to give a fuller perspective of the data.

4. FAIRNESS

To explore the concept of fairness, we use a hypothetical directed to the question of fairness in entering job placement. The hypothetical is simplified for ease of arithmetic, but it reflects salient features of a group of employees at the Harris Bank who were hired between 1969 and 1971 (Roberts 1979). The income variable is categorical, entering job at time of hire. It is assumed that there are two homogeneous entering job groups, clerical and professional, and that there are three levels of education at entry: high school, college, and MBA. Professional jobs are deemed more desirable than clerical, and education is assumed to be a valid, job-related qualification. The joint frequency distribution of the data appears in Table 4.

Table 4. Hypothetical Distribution of Male and Female Employees by Education and Entering Job

Education	Male			Female		
	Clerical	Professional	Total	Clerical	Professional	Total
High School	40	0	40	80	0	80
College	23	17	40	16	4	20
MBA	1	19	20	0	0	0
Total	64	36	100	96	4	100

The bottom margin of Table 4 shows that 96 percent of the females, compared with 64 percent of the males, are in clerical positions. The right column of each subtable shows that 80 percent of the females, compared with 40 percent of the males, have only a high school education. Both comparisons are relevant to an understanding of fairness; one cannot simply look at one and exclude the other.

In this simple example, regression entails study of the two sets of conditional distributions corresponding to the two directions of percentaging the joint frequency distribution. Horizontal percentaging corresponds to the conditional distributions of entering job given education; this is direct regression. Vertical percentaging corresponds to the conditional distributions of education given entering job; this is reverse regression.

Examine now Tables 5 and 6. When interpreting the tables, we must keep in mind that the females on average have lower educational levels than do the males and are concentrated in lower job positions. Table 5 presents the relation between education and entering job in the traditional regression perspective used in discrimination studies. All high school graduates, male or female, are placed in clerical jobs, and, given the reasonable assumption that clerical jobs demand less educational background, this placement appears to be fair. Virtually all male MBA's receive professional jobs, but since there are no females with MBA's, the given data do not provide an assessment of fairness for this group. (Possible unfairness may enter in the search for job applicants with an MBA degree or in the selection among them for initial hire. Additional data would be required to investigate these questions.)

Of the 40 male college graduates, 42.5 percent are placed in professional jobs, while of the 20 female college graduates, only 20.0 percent are placed in professional

Table 5. Conditional Distribution of Entering Job Given Education for Male and Female Employees in Table 4—Type 1 Unfairness

Education	Male			Female		
	Clerical	Profes-sional	Total	Clerical	Profes-sional	Total
High School	100.0	.0	100.0	100.0	.0	100.0
College	57.5	42.5	100.0	80.0	20.0	100.0
MBA	5.0	95.0	100.0	.0	.0	.0
Total	64.0	36.0	100.0	96.0	4.0	100.0

Table 6. Conditional Distribution of Education Given Entering Job for Male and Female Employees in Table 4—Type 2 Unfairness

Education	Male			Female		
	Clerical	Profes-sional	Total	Clerical	Profes-sional	Total
High School	62.5	.0	40.0	83.3	.0	80.0
College	35.9	47.2	40.0	16.7	100.0	20.0
MBA	1.6	52.8	20.0	.0	.0	.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

jobs. (For the purpose of this hypothetical, and throughout the article, we abstract from the question of statistical significance of differences.) Assume that the entering job placement is entirely the employer's decision, and that all new employees aspire to professional jobs. (Both assumptions are challengable in applications.) Unless their college degrees are less job-related than those of men, it appears that the females are treated unfairly. We have identified groups of males and females with the same job qualifications, and the males more often get the higher job. ("Females have to be better qualified to attain higher job levels.")

From Table 4, we see that there are 64 males and 96 females in clerical jobs. Now examine Table 6. The distribution of educational level in these clerical jobs is higher for males than for females. Similarly, there are 36 males and 4 females in professional jobs, and the distribution of educational level within professional jobs also is higher for males than for females. Within either job group the males tend to have higher qualifications than do their female coworkers. Unless there is special information about differences in jobs or degrees, it appears that the males are treated unfairly. We have identified groups of males and females with the same jobs, and the males more often have the higher job qualifications. ("Males have to be better qualified to hold current job levels.")

4.1 Two Types of Fairness

We name and define two types of fairness (or unfairness):

1. Fairness 1: the conditional distributions of income given job qualifications are the same for both sexes.
2. Fairness 2: the conditional distributions of job qualifications given income are the same for both sexes.

With additional information about individual differences in background, training, and other job qualifications, the picture relevant to the two types of fairness can change. The definitions are to be taken with respect to some given set of qualifications, usually those available for statistical study. Almost all discussion of unfairness has centered on unfairness 1. We believe this to be a consequence of statistical habits that focus attention on income as a dependent variable.

In the hypothetical, females suffer from unfairness 1; males, from unfairness 2. Unfortunately, as illustrated here, one cannot always (or even usually) hope for both types of fairness. Two special cases when the concepts coincide are the following: (a) Identical joint distributions of qualifications and income for males and for females, (b) An exact relationship between income and qualifications, with identical regressions for the two sexes.

It would be possible to remove unfairness 1 to females in the hypothetical example by the following modification of Table 4: place 14 of 40 college males and 7 of 20 college females in professional jobs, and keep the number of clerical and professional jobs fixed. Table 7 shows the result of this modification.

If one computes the vertical percentages, one can see that unfairness 2 against males in both occupational groups has increased. Moreover, placing all the college females in professional positions, with compensating changes for college males, would result in unfairness 1 as well as unfairness 2 against males. And what is more surprising, even the most extreme unfairness 1 to females would not remove unfairness 2 to males.

The reason that the hypothetical leads to surprising conclusions is that, on average, the males have higher job qualifications. There is empirical evidence to suggest that the hypothetical is realistic in many applications. For example, evidence from direct regression studies of salary discrimination often supports the conclusion that female shortfalls in job qualifications tend to be found when female shortfalls in salaries are found. Statistical habits, however, make us notice the female shortfalls in income, while we do not notice the female shortfalls in job qualifications. This tendency seems to apply equally to statisticians and to nonstatisticians.

4.2 Appearance of Discrimination

The two concepts of fairness will sometimes be in agreement and point to both unfairness 1 and 2 against

Table 7. Modified Distribution of Male and Female Employees by Education and Entering Job—Fairness 1

Education	Male			Female		
	Clerical	Profes-sional	Total	Clerical	Profes-sional	Total
High School	40	0	40	80	0	80
College	26	14	40	13	7	20
MBA	1	19	20	0	0	0
Total	67	33	100	93	7	100

a certain sex group. Table 8 gives a hypothetical that illustrates this situation. Computation of the vertical percentages shows that 27 percent of the females in clerical jobs have a college education compared with 20 percent of the males. Also, 80 percent of the females in professional jobs have a college education whereas the corresponding percentage for males is 73 percent. These percentages point to unfairness 2 against females.

Turning to the proportions of females in professional jobs given level of education, we see that 8 percent of the females with high school degrees are in professional positions. However, 50 percent of the males with the same qualification are in professional positions. Also, 50 percent of the females with a college education are in professional positions as opposed to 92 percent of the males. Clearly, this shows unfairness 1 against females.

Based on the given data, females experience both types of unfairness. It is conceivable that information about additional job qualifications would change the picture. For example, the males may have tended to specialize in academic subjects closely related to the company's requirements for professional jobholders.

But there is always a point at which no additional qualifications are available and one must infer the presence or absence of unfair treatment based on the evidence at hand. In the past, these decisions have been triggered largely by the concept of type 1 unfairness. Type 2 unfairness, problems of economic definition of discrimination, and consideration of underlying causal structure have received little attention.

One final note on fairness is important. The definitions of fairness implicitly assume both that measured job qualifications are available and that the available qualifications are related to salary or job level for males and females separately. If no measured qualifications are available—there is literally no information in the data on individual differences in ability, training, experience, and so on—then there is only one type of unfairness: the group with the lower mean salary or job level is unfairly treated. This situation might be applicable to certain classes of unskilled jobs. If qualifications are independent of job level or salary for both males and females, then, as we argue below in a specific example, the job qualifications should be dropped from consideration.

4.3 Generality of the Hypothetical

The ideas of fairness 1 and fairness 2 have been presented in the context of simple cross tabulations. All the concepts carry over to the more elaborate salary regressions often done in discrimination studies. Whenever the relationship between income and qualifications is imperfect, one cannot simultaneously have fairness 1 and fairness 2, unless the joint distributions are identical for males and females (Dempster 1979).

It is possible to have both unfairness 1 and unfairness 2 against the same group. Then, the two types of fairness point to the same conclusion about unfair treatment. This conclusion could be reversed by additional data or by the injection of judgments that go beyond the data available.

But often the two types of fairness will suggest contradictory conclusions about unfair treatment. Such contradictions pose serious problems for lawyers, statisticians, and judges alike. They cannot be dispelled by loose rhetoric.

5. CRITICISMS OF REVERSE REGRESSION

The ideas of reverse regression have attracted interest and criticism since they were applied to the Harris Bank case (Roberts 1979). A number of comments, oral and written, published and unpublished, have come to our attention. Many of the criticisms turn on questions of fairness, as developed in Section 4, rather than technical questions of statistics. In particular, they appear to resist the relevance of fairness 2, which, in turn, is assessed by reverse regression. In this section, we illustrate those criticisms and discuss the issues with reference to both types of fairness.

5.1 Finkelstein

Finkelstein (1980) argues that reverse regression fails to deal with possible discrimination in promotion. He offers two interesting hypotheticals to convey his reasoning. For ease of comparison with our own hypotheticals in Section 4, we change his terms "low-paying" and "high-paying" to "clerical" and "professional"; and his terms "other" and "highly qualified" to "high school" and "college."

Table 9 gives Finkelstein's first hypothetical, in which he considers 45 male and 45 female employees before and after a mass promotion. We shall interpret Finkel-

Table 8. Hypothetical Distribution of Male and Female Employees by Education and Job Illustrating Unfairness Against Females

Education	Male: Clerical	Male: Professional	Male: Total	Female: Clerical	Female: Professional	Female: Total
High School	20	20	40	55	5	60
College	5	55	60	20	20	40
Total	25	75	100	75	25	100

Table 9. Distribution of Male and Female Employees by Education and Job—First Finkelstein Hypothetical

Education	Male: Clerical	Male: Professional	Male: Total	Female: Clerical	Female: Professional	Female: Total
a. Before Promotion						
High School	20	0	20	25	0	25
College	25	0	25	20	0	20
Total	45	0	45	45	0	45
b. After Promotion						
High School	10	10	20	20	5	25
College	5	20	25	10	10	20
Total	15	30	45	30	15	45

stein's hypothetical in the language of fairness. Before promotion, one can discuss only fairness 2, since fairness 1 is not defined in terms of a single job group. Because the males include a higher percentage of college graduates, they are in an unfair initial position. Promotion, on the other hand, is unfair to females in the type 1 sense, since for both levels of education, the percentage of females promoted is less than for the corresponding male group. The final position, however, demonstrates fairness 2, since the percentage of college graduates within each job group is the same for males and females.

Finkelstein's concern is that an unfair (sense 1) promotion leads to a fair (sense 2) employment position at the end. Finkelstein does not seem to point out that the initial position was unfair in sense 2. Our interpretation is that if one considers fairness 2, unfair promotion corrects an unfair initial position. This reasoning parallels the familiar civil rights position that correction of past discrimination may require differential treatment in favor of those discriminated against.

The fairness or unfairness of the picture before promotion cannot be assessed within the framework of direct regression because fairness 1 is not defined in terms of a single job group.

Finkelstein's second hypothetical is a variation on the first. The promotion rates, given education, are the same for both sexes, but the final situation is unfair to the males in sense 2. Table 10 gives details. In this hypothetical, the initial employment position is unfair (sense 2) to males, more so than in the previous hypothetical. The fair (sense 1) promotions necessarily leave males in an unfair (sense 2) position afterwards in both job groups.

5.2 Weisberg and Tomberlin

Weisberg and Tomberlin (1982) present a statistical definition of discrimination that is stated in terms of fairness 1. Although they discuss Dempster's treatment of what we call fairness 2, they see no room for this concept in the definition of discrimination, at least under existing law. Their analysis of reverse regression resembles that of Finkelstein, except that they make explicit the difference in the starting distributions of male and female qualifications. This difference, indeed, is central to their criticism of reverse regression. Weisberg and Tomberlin argue that these differences may have existed prior to any action taken by the employer and it is natural to condition on them. Since reverse regression does not condition on preexisting job qualifications, they conclude that the logic of reverse regression is faulty. Because Weisberg and Tomberlin's definition of discrimination only permits consideration of fairness 1, it is understandable that they would find fairness 2 incompatible with their definition of nondiscrimination.

Table 10. Distribution of Male and Female Employees by Education and Job—Second Finkelstein Hypothetical

Education	Male: Clerical	Male: Professional	Male: Total	Female: Clerical	Female: Professional	Female: Total
a. Before Promotion						
High School	15	0	15	30	0	30
College	30	0	30	15	0	15
Total	45	0	45	45	0	45
b. After Promotion						
High School	10	5	15	20	10	30
College	10	20	30	5	10	15
Total	20	25	45	25	20	45

Weisberg and Tomberlin offer an interesting hypothetical in support of their criticisms of reverse regression.

> ... suppose a company wishes to select six individuals for a particular position. There are ten males and ten females available. They are all scored on a test that has been validated as a perfect predictor of future productivity, with the following results:
>
> Males: 40, 50, 60, 60, 70, 75, 80, 90, 95, 95
> Females: 30, 35, 55, 60, 60, 75, 75, 75, 85, 90
>
> Then, if the employer selects on the basis of this index, he will choose the top four men and the top two women to obtain
>
> Selected Group: 80(M), 85(F), 90(M), 90(F), 95(M), 95(M)
>
> Note that although the requirements for men and women are identical, the distribution of qualifications will differ between the males and females selected. In particular, the average level of Q [productivity] for males selected is 90.0, while that for women is 87.5.
>
> Does this represent discrimination against males, who seem after the fact to have had the more stringent standards applied? Clearly not. It simply represents the application of a non-discriminatory rule where the marginal distributions of the two groups differ. The conditional probability of Q given S [salary] is determined not only by the employer's requirements, but also by the pre-existing population distribution for the groups.

Here true productivity is assumed to be known to all: the authors of the hypothetical, the employer, and any statistician who comes along to audit the employer's performance. As we shall develop briefly in Section 6, the assumption of known productivity removes any unfairness suggested here for all who accept the economic proposition that salary should equal the value of the marginal product. Because true productivity is known, the income variable, selection for the job, should match the true productivity variable, score on the test. An inconsistency arises because selection for the job is a discrete variable whereas the score on the test is continuous.

There are two ways of resolving the apparent inconsistency. First, it can be argued that the measurement reflecting true productivity is not the score on the test, but rather a threshold variable that is derived from the score. In this example, the pertinent qualification is whether the employee scores among the highest six scores in order to be selected for the job. The threshold score variable might be defined to assume the value 1 if the employee scores at or above 80 and the value 0 otherwise. This variable is then a perfect predictor of job selection so that both fairness 1 and 2 prevail.

For the second argument, assume that the actual score on the test reflects true productivity. It appears that equal salaries for the six selected employees are assumed and that this is discriminatory. Suppose that the specified salary is 80 units. Then the employer is discriminating against five of the six employees. In aggregate, the males are underpaid by 10+15+15 = 40 units, 10 units on average; the females are underpaid by 5+10 = 15 units, or 7.5 units on average.

The introduction of salary into the hypothetical reminds us of the multidimensional nature of "income." Each of the six employees gets both a job and a salary. Because productivity is known, the employer can achieve fairness 2 by adjusting salaries. In particular, salaries can be equated to the known productivities. If there is some constraint that requires equal salaries, the employer can adjust the job content to fit the added qualifications of the more qualified individuals. Unfairness 2 can always be removed by adjusting salary or job content, and it seems both fair and rational for the employer to do so.

5.3 Blattenberger and Michelson

Blattenberger and Michelson (1980) expressed several criticisms of reverse regression, most of which are anticipated by the exposition in the earlier part of this article. One major exception, which was also presented in Michelson's testimony at the Harris Bank hearing in 1979, concerns the role of productivity in discrimination cases. For completeness of coverage, we deal briefly with this point.

Michelson questioned the extent to which employers try to equate individual employee incomes to productivity, rather than to follow some other criterion such as formula payment for measured job qualifications. For example, suppose that an organization computes all salary increases strictly as a function of seniority and years of formal education. Such an organization cannot discriminate by sex in its wage increases. (There would remain, however, the interesting question of possible age discrimination.)

If employers strictly follow a formula payment approach, then there is a perfect correlation between pay and job qualifications as embodied in the formula. However, the formula itself may or may not be accepted as fair. Moreover, to the extent that employers try to make some judgmental modifications of the results of a computational formula based on measured job qualifications, it is hard to see how the idea of productivity can be left out of consideration. Fairness 2 is then relevant.

5.4 Discrimination by Tokenism

An interesting criticism can be conveyed by the following hypothetical. In this hypothetical 100 males and 100 females have identical distributions of education; 140 clerical and 60 professional jobs are to be filled. Assume for argument that an employer wants to discriminate against females, knowing that the company is later going to be audited by a statistician using reverse regression. Is it possible to cheat if the judgment is to be based solely on fairness 2? Table 11 presents a proposed strategy for cheating. The employer is fair in sense 2, and appears to pass the test of reverse regression, although he manifestly fails on direct regression and thus violates fairness 1.

Table 11. Hypothetical Distribution of Male and Female Employees by Education and Job—Discrimination Using Tokenism

Education	Male			Female		
	Clerical	Professional	Total	Clerical	Professional	Total
High School	25	25	50	45	5	50
College	25	25	50	45	5	50
Total	50	50	100	90	10	100

At this point, we return to an assumption about both fairness 1 and fairness 2 made in Section 4. This is the requirement that job qualifications be correlated with job level or salary for males and females separately. In both direct and reverse regression, comparisons of statistically adjusted incomes, or of indices of qualifications, between males and females must be based on the joint distribution of income and job qualifications within each of the two sex groups. If within each group, income and job qualification are independent, as in this case, neither set of conditional distributions is informative. In Table 11, the within-group joint distributions suggest either that the education variable is not a meaningful job proxy, or that job position is not a meaningful income variable. If the second possibility is ruled out, then the first remains; which suggests that no regression adjustment, direct or indirect, is in order. That leaves us with Table 12.

In the absence of further mitigating evidence from other job qualifications, Table 12 suggests prima facie evidence of discrimination against females. For, without job qualifications to work with, statistical comparison must stop with the comparison we termed "naive" in Section 3. The discriminating employer has painted himself into a corner by creating a situation in which the employee's education level is of no use in his defense.

A more extreme example of the strategy illustrated in Table 11 is shown in Table 13. Again, the statistical

Table 12. Joint Distribution of Education and Job for Male and Female Employees in Table 11

Job Group	Male	Female
Clerical	50	90
Professional	50	10
Total	100	100

Table 13. Hypothetical Distribution of Male and Female Employees by Education and Job—Discrimination by Exclusion

Education	Male			Female		
	Clerical	Professional	Total	Clerical	Professional	Total
High School	20	30	50	50	0	50
College	20	30	50	50	0	50
Total	40	60	100	100	0	100

defense based on education has been lost. The complete absence of female professional employees creates a strong suspicion of discriminatory collusion against females, to be investigated by direct search for evidence.

A variation on Table 11 entails a hypothetical in which there are only college graduates and the high school line is deleted. Then education contributes nothing to the prediction of income, and, again, the employer has no defense against the prima facie case of discrimination. We have examined more elaborate variations in which one introduces some weak correlation between qualifications and income. The variations immediately lead to unfairness in both senses 1 and 2. But further explorations are needed to see if there is in fact a subtle way of cheating through tokenism.

5.5 Ferber and Green

We have seen a deep potential conflict between concepts of fairness 1 and fairness 2. In actual data, one must often decide which concept is to be accorded the principal weight. We see no simple principle to guide this decision, at least as long as attention is confined to the available data and additional, subjective information is not introduced.

Ferber and Green (1982) suggest a general presumption in favor of fairness 1. "Upon careful consideration it appears that [fairness 1] . . . is fair to people . . . [fairness 2] is fair to jobs . . . ." We do not agree. Both types of unfairness are unfairness to people. Consider the following introspection.

Suppose that employee A has exactly the same qualifications as employee D and finds that D is making $5000 a year more. Employee A feels—rightly, on the evidence assumed—the victim of unfairness 1.

Suppose now that employee A finds that employee R has the same salary, but is substantially less qualified. Employee A feels, rightly, the victim of unfairness 2. Perhaps employee A would feel worse about one or the other comparison, but in no sense is one comparison personal and the other impersonal.

6. BEYOND FAIRNESS: ECONOMICS AND ECONOMETRICS

Direct and reverse regression comparisons for assessing fairness can be extended to economic concepts of discrimination. In competitive markets, wages are determined by forces labeled "supply" and "demand." In a deterministic world, economic nondiscrimination implies that the actual wage equals the value of the marginal product of labor (see, e.g., Becker 1957). For brevity, we use the popular term "productivity" in place of value of the marginal product. Productivity, denoted by P, reflects employee skill characteristics and other factors that affect the output or actual job performance of the employee. In the interest of maximizing profits, employers link compensation to productivity.

Few statisticians in discrimination studies have attempted to estimate productivity directly from output data. (The main exceptions are studies of college faculties, for which limited output measures, such as papers published and teaching ratings, are available.) In many studies, the statistician tries to learn about productivity by modeling relationships between salaries and employee job qualifications.

Suppose that an auditing statistician knows true productivity P for all employees of a firm and wants to investigate issues of fairness with regard to salary, Y. The statistician could regress Y on P and *Sex* (paralleling direct regression) and P on Y and *Sex* (paralleling reverse regression). As long as P and Y are imperfectly correlated and the joint distribution of P and Y differs between males and females, the statistician would be faced with the exact counterpart in economic terms (productivity) of our analysis of fairness. For example, the findings might be

1. Females have lower mean productivity than males.
2. For given productivity, females have lower mean salaries than males.
3. For given salary, male and female mean productivity are the same.

In this example "discrimination 1" against females would parallel "unfairness 1" against females, and "nondiscrimination 2" would parallel "fairness 2."

Of course, there is much more to statistical study of employment discrimination than fairness alone. True productivity is rarely known. It may be estimated using a model of the firm involving a production function. The use of job qualifications as proxies for productivity is almost a choice of desperation, dictated by limitations of available data. In most realistic applications, the qualification list is far short of what is needed for a precise assessment of productivity. This limitation poses serious problems regarding statistical inference from either direct or reverse regression models.

Dempster (1981) argues that statistical inference about discrimination must rely on underlying causal mechanisms. Such inferences depend both on the data and on prior beliefs about the real processes underlying the data. The theory of labor economics has much to contribute to an understanding of how economic factors affect the employer's decisions on hiring, termination, promotion, and compensation of workers, as well as the employee's decisions to accept a job, seek promotion, or demand a wage increase. Economic concepts are essential for formulating structural models of the employment process that are used in discrimination studies.

In studies of employment discrimination, a number of econometric and statistical questions emerge that pertain to proper model specification. For example, in order that job qualifications (X) and income (Y) be linearly related, certain conditions on the joint distribution of X and Y must be satisfied. Pronounced nonlinearity or interaction effects require the specification of nonlinear models. When linear regression models are appropriate, the presence of measurement errors in the variables can lead to biased estimates of the regression coefficients (Madansky 1959). A number of authors have proposed methods for testing whether stochastic regressors are independent of the disturbances in regression models (see, e.g., Wu 1973 and Reynolds 1982). Questions that pertain to the inclusion of specific X variables in the model and the effects of omitted variables must also be addressed.

Recent work in econometrics points out potential biases that occur in salary or placement models owing to selection rules for hiring employees (Heckman 1979). Employment studies of discrimination often involve a separate analysis of the applicant pool to determine if there are biases in the hiring process that affect subsequent placement and salary. Abowd and Killingsworth (1981) propose an econometric formulation of both the selection and wage offer decisions in a simultaneous equation framework. We hope to develop, in a future article, an analysis of some of these issues that have been raised in connection with reverse regression.

Economic views of discrimination and econometric models of the employment process are still in an evolutionary stage of development. New statistical methods for assessing discrimination have emerged from the need for these assessments in legal cases. These methods are also evolving and becoming more sophisticated as the economic and legal aspects of discrimination become better understood.

Our main conclusion is that fairness is central to interpretation of data on discrimination. Both direct and reverse regression are valuable tools for determining what the data have to say about fairness.

ACKNOWLEDGMENTS

The authors are indebted to many colleagues and coworkers for stimulating discussions; to Jeffrey J. Keenan for the computations in Section 2; and to Professor Arthur P. Dempster, who first made us aware of the concepts that we have called "fairness 1" and "fairness 2" and who has helped us to better understand issues arising in studies of employment discrimination. The authors benefited from helpful comments by the editor and a referee.

[Received June 1982. Revised September 1982.]

REFERENCES

ABOWD, JOHN M., and KILLINGSWORTH, MARK R. (1981), "An Analysis of Hispanic Employment, Earnings and Wages in the Federal and Non-Federal Sectors with Special Reference to Puerto Ricans," U.S. Department of Labor, Report Number DOL ETA 21-36-78-61.

BECKER, GARY S. (1957), *The Economics of Discrimination in Employment* (2nd ed.), Chicago: University of Chicago Press.

BIRNBAUM, MICHAEL H. (1979a), "Is There Sex Bias in Salaries of Psychologists?" *American Psychologist*, 34, 719–720.

——— (1979b), "Procedures for the Detection and Correction of Salary Inequities," in *Salary Equity: Detecting Sex Bias in Salaries Among College and University Professors*, eds. Thomas R. Pezzulo and Barbara E. Brittingham, Lexington, Mass.: D.C. Heath, Ch. 10.

BLATTENBERGER, GAIL, and MICHELSON, STEPHEN (1980), "Reverse Regression Analysis of Wage Discrimination: A Critique," unpublished paper, Econometric Research Inc., Washington, D.C.

DEMPSTER, ARTHUR P. (1979), "Statistical Concepts of Discrimination," unpublished paper, Harvard University, Dept. of Statistics.

——— (1981), "Causal Inference, Prior Knowledge, and the Statistics of Employment Discrimination," unpublished paper, Harvard University, Dept. of Statistics.

FERBER, MARIANNE A., and GREEN, CAROLE A. (1982), "Employment Discrimination: Reverse Regression or Reverse Logic," unpublished paper, University of Illinois, Champaign.

FINKELSTEIN, MICHAEL O. (1980), "The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases," *Columbia Law Review*, 80, 737–754.

FISHER, FRANKLIN (1980), "Multiple Regression in Legal Proceedings," *Columbia Law Review*, 80, 702–736.

GWARTNEY, J., ASHER, E., HAWORTH, C., and HAWORTH, J. (1979), "Statistics, the Law and Title VII: An Economist's View," *Notre Dame Lawyer*, 54, 633–660.

HECKMAN, JAMES J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

MADANSKY, ALBERT (1959), "The Fitting of Straight Lines When Both Variables are Subject to Measurement Error," *Journal of the American Statistical Association*, 54, 173–205.

REYNOLDS, ROGER A. (1982), "Posterior Odds for the Hypothesis of Independence Between Stochastic Regressors and Disturbances," *International Economic Review*, 23, 479–490.

ROBERTS, HARRY V. (1979), "Harris Trust and Savings Bank: An Analysis of Employee Compensation," Report 7946, CMSBE, Graduate School of Business, University of Chicago.

——— (1980), "Statistical Biases in the Measurement of Employment Discrimination," in *Comparable Worth: Issues and Alternatives*, ed. E. Robert Livernash, Washington, D.C.: Equal Employment Advisory Council.

SCHLEI, B., and GROSSMAN, P. (1976), *Employment Discrimination Law*, Washington: B.N.A., Chapter 36.

WEISBERG, HERBERT I., and TOMBERLIN, THOMAS J. (1982), "Statistical Evidence in Employment Discrimination Cases," *Sociological Methods and Research*, in press.

WU, DE-MIN (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41, 733–750.