proxy variables on regression coefficients. Graduate School of Business Administration, Harvard Univ.

MEIER, P. (1986). Damned liars and expert witnesses. *J. Amer. Statist. Assoc.* **81** 269–276.

PHELPS, E. S. (1972). The statistical theory of racism and sexism. *Amer. Econ. Rev.* **62** 59–61.

PRATT, J. W. (1986). Review of "W. G. Cochran's Impact on Statistics." *J. Amer. Statist. Assoc.* **81** 565–566.

PRATT, J. W. and SCHLAIFER, R. (1984). On the nature and discovery of structure. *J. Amer. Statist. Assoc.* **79** 9–21.

SAVAGE, L. J. (1962). *The Foundations of Statistical Inference.* Methuen, London.

SAVAGE, L. J. (1967). Implications of personal probability for induction. *J. Phil.* **64** 593–607. Reprinted in *The Writings of Leonard Jimmie Savage—A Memorial Selection*, published by ASA and IMS.

SHAFER, G. (1976). *A Mathematical Theory of Evidence.* Princeton Univ. Press, Princeton, N. J.

SHAFER, G. (1987). Probability judgment in artificial intelligence and expert systems. *Statist. Sci.* **2** 3–16.

Comment

Franklin M. Fisher

Arthur Dempster's paper has a good deal to say about the interpretation of probability models and causal thinking, much of it uncontroversial. Rather than discuss such matters in the abstract, however, let's consider the example of employment discrimination that Dempster uses and see what it is that he is really saying.

This is not hard for me to do, because I have encountered Dempster's views on previous occasions. I was a witness for the plaintiff in two employment discrimination cases, *OFCCP v. Harris Trust and Savings Bank* (Department of Labor Case No. 78-OFCCP-2) and *Cynthia Baran v. The Register Publishing Company* (Civil N. 75-272, U. S. District of Conn.). In both cases, I testified on matters of econometric and statistical principle rather than putting forward a study of my own, and Dempster testified for the defendant. This paper is largely based on my experience and testimony in those cases. (I believe—but do not know for sure—that, just as my own experience in employment discrimination cases has been as an expert assisting plaintiffs' counsel, Dempster's experience, to which he refers, has been as an expert assisting counsel for defendants.)

A particular employer is accused of sex discrimination. (As does Dempster, I take this as a leading example.) In general, this means that salaries paid to female employees average less than those paid to male employees. One possible reason for this discrepancy is discrimination; another is that male employees are more productive than female ones.

To examine the question of whether there is a gender-based wage difference holding productivity constant, a statistician estimates the model

$$Y = G\alpha + X\beta + e, \tag{1}$$

where (letting i denote values for a particular employee), Y_i denotes salary, G_i is 0 for female and 1 for male employees, X_i is a vector of observed employee characteristics (education, experience, age, etc.), and the e_i are assumed to be random variables (usually taken to be independent $N(0, \sigma^2)$, although this will play no role in the present paper). α, β and σ are parameters to be estimated, and it will aid discussion to assume that the sample size is sufficiently large to enable us to take such parameters as known with certainty. A positive value of α is taken to be evidence of discrimination against females.

What is wrong with such a procedure? Dempster points out several possibilities. In the first place, he suggests interpreting the stochastic element involved by assuming that the nondiscriminatory employer is computing

$$Y^* = E(Y^{**} \mid G, X^*), \tag{2}$$

where X_i^* is a vector of employee characteristics known to the employer (but possibly not to the analyst), Y_i^{**} denotes "true" employee productivity and Y_i^* denotes employee productivity as estimated by the employer in (2). Both Y_i^{**} and Y_i^* are assumed measured in monetary units to be comparable to wages, Y_i. Discrimination is to be interpreted as paying males more than Y_i^*, i.e.,

$$Y = G\alpha' + Y^*, \tag{3}$$

with $\alpha' > 0$.

This is not the only form that discrimination can take. Depending on the state of the outside labor market, discrimination is more likely to consist of paying females *less* than the employer truly thinks

Franklin M. Fisher is Professor of Economics, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

they are worth than of paying them what their estimated productivity indicates and paying males more. To alter Dempster's treatment in this regard, however, would make no essential difference to either his or my discussion.

In any event, if (2) is assumed to take a linear form, it becomes

$$Y^{**} = G\alpha'' + X^*\beta^* + e^{**}, \tag{4}$$

so that wages are determined in terms of characteristics observable to the employer by

$$Y = G\alpha^* + X^*\beta^* + e^{**}, \tag{5}$$

and discrimination against females means $\alpha^* > \alpha''$. (The fact that Dempster allows α'' to be positive is an issue I discuss later. Until further notice, assume it to be zero.)

Now, (5) looks very much like (1). Indeed, if the statistician and the employer had the same information, X^*, they would be the same. They differ because the employer is assumed to have some information not available to the statistician. Hence, it is possible that estimation of (1) will lead to a positive α and therefore to a finding of discrimination even where no discrimination is present.

We come now (not for the last time) to the question of what arguments are to be taken seriously. Every employer accused of sex discrimination will certainly argue that he (given the example it is unlikely to be "she") can judge productivity better than the statistician can and that a regression based on the information that he, the employer, actually uses would show no discrimination.

Such an argument may, indeed, be true. But if this is all the employer says, if he proffers no description of the inside information he possesses and offers no positive evidence as to its use, then acceptance of that argument is tantamount to finding for the employer in *all* cases. For this has nothing intrinsically to do with statistics. Against *any* evidence of sex discrimination short of a "smoking gun" memo, the employer can always reply: "It only looks like discrimination, but if you only had my secret (and unrevealable) information on productivity you would see that it is not."

It is for this reason that a finding of sex discrimination based on (1) is sometimes taken as establishing a *prima facie* case. The defendant ought to be required to put forward an affirmative showing that such a finding is wrong rather than permitted to rebut it on the basis of undisclosed productivity information.

There are two forms that such an affirmative showing can take. The strongest is a regression analysis using the employer's full information, X^*. Indeed, one might argue that if such information really plays a systematic and important role in wage decisions, then the employer can be expected to have retained or be able to reconstruct it so that it can be used. On this view, the very absence of such information makes the employer's argument suspect.

On the other hand, such a view may be too stringent. Wage decisions are not in fact made using regressions; further, data may be impressionistic and not retained. Hence, while one may be suspicious of an employer who makes the secret information argument without being able to produce the information, there may be something more to be said.

A less stringent standard is to permit the employer to argue that (1) yields a biased estimate of α^* based on a *specific* description of the information he claims to use and the way in which the lack of that information operates if (1) is used. This is at least an intellectually respectable argument (and one which Dempster uses). Even so, because such arguments tend to have testable implications, I would tend to treat them skeptically if the employer produces no empirical justification. Employment discrimination cases can involve a great deal of money—enough to justify at least sampling the employees to recover the information involved.

Put this aside, however, and consider directly the argument that unincluded information can produce a positive α in (1) when α^* is in fact zero. When will this happen?

For simplicity, suppose first that X and X^* are both single variables (suppressing any constant term in (4) and (5)), and that the information problem consists of the fact that X is a noisy measure of X^*. That is,

$$X = X^* + v, \tag{6}$$

where v is uncorrelated with the "true" variable, X^*, and also uncorrelated with G. For example, if X is the score on a particular aptitude test, X may be influenced not only by the employee's true aptitude (somehow observable to the employer through other means) but also by the state of health of the employee on the day the test was taken.

This is a classic errors-in-variables problem. We know that if X is used instead of X^* in (5), its estimated coefficient (assumed positive) will be biased toward zero. Because the very fact that (1) was estimated rather than relying on the raw comparison of male and female wages means that male employees are likely to have scored higher than female ones in terms of the variable used as X, G is likely to be positively correlated with X. This will produce an estimate of α^* that is biased upward. In common sense terms, the use of X rather than X^* will underestimate the true effects of X^* on wages. This, in turn, will lead to too small a correction of the male-female wage differential for productivity differences and hence to an overestimation of the effects of gender. (The matter is less simple if there is more than one variable in X or X^*, but there will be some tendency in this direction in any case.)

For my purposes, the errors in variables explanation of the source of the problem is slightly inconvenient. An equivalent way of proceeding is to substitute (6) into (5), obtaining

$$Y = G\alpha^* + X\beta^* + (e^{**} - v\beta^*). \tag{7}$$

With $\beta^* > 0$ and $(-v\beta^*)$ part of the error term, X is negatively correlated with the error, leading to a downward bias in its estimated coefficient (and an upward bias in that of G, as before).

Harry Roberts first proposed "reverse regression" to deal with this problem. Dempster is incorrect when he states, "The original motivation for $\cdots$ reverse regression, as well as for the contrasting terms reverse and direct, comes from contrasting definitions of 'fairness' $\cdots$." The use of reverse regression in that regard was a later development, apparently partly suggested by Dempster himself. (See Roberts (1979, 1980), Ferber and Green (1984, page 111) and especially, Conway and Roberts (1983, page 85). I discuss the "fairness" concepts below.)

Reverse regression deals with the errors in variables problem in the following way. Suppose that (7) is solved for X, obtaining

$$\begin{aligned} X = {} & Y(1/\beta^*) + G(-\alpha^*/\beta^*) \\ & + \{e^{**}(-1/\beta^*) + v\}. \end{aligned} \tag{8}$$

If this is estimated by regression, the presence of v causes no difficulty because, by assumption, it is uncorrelated with G and Y. This suggests the estimation of (8)—reverse regression—rather than (7) to obtain an estimate of α^*.

Of course, once it is written out in this way, one of the problems with reverse regression becomes obvious. Although the presence of v in the disturbance term in (8) creates no difficulty, the presence of e^{**} certainly does. Assuming that the firm used (4) and (5) to set wages, Y is certainly positively correlated with e^{**} and hence negatively correlated with the disturbance term in (8). This means that estimating (8) by regression will lead to an estimate of $(1/\beta^*)$ that is biased downward and hence to an implied estimate of β^* itself that is biased upward. Hence, reverse regression will *over*state the effects of X^* on productivity, leading (by the same reasoning as before) to an *under*estimate of α^*, the all important coefficient of G in (5).

Note that this means that a finding that reverse regression leads to a lower estimate of α^* than does direct regression (or even a finding that reverse regression shows no sex discrimination whereas direct regression does) does not imply that the errors in variables problem is a serious one for direct regression.

I shall return to the serious problem for reverse regression caused by the presence of e^{**} in the disturbance term of (8). For the present, however, put it to one side, for there are other problems to consider.

The first of these is technical. Once we drop the assumption that X and X^* are single variables, the analysis becomes more complex. For one thing, it is no longer guaranteed that the errors-in-variables problem will lead direct regression to an estimate of α^* that is biased upward (although this seems a likely outcome). More important, it is far from clear what reverse regression is to mean. The natural composite is $X^*\beta^*$, but the lack of a consistent estimate of β^* (let alone information on X^*) is what causes the problem in the first place.

There is a more fundamental problem than this, however. Whether or not X and X^* are single variables, the assumption that leads to the errors-in-variables analysis is unlikely to be satisfied. In terms of (6), this is the assumption that v, the difference between X and X^*, is correlated with X, but not with X^*. As soon as we leave the example of aptitude test scores, that assumption stops being plausible. In practice, the employee characteristics available to the statistician are generally such things as years of education, age, experience and so forth. Although it is certainly possible that the employer observes other attributes that contribute to productivity, this is likely to mean that

$$X^* = XH + W, \tag{9}$$

where H is a matrix of parameters and W is orthogonal to X, not to X^*. In other words, the assumption that the employer has more information than the statistician is likely to mean just that. The problem is not that such variables as education, age and experience are noisy substitutes for some "true" productivity measure that the employer observes, but rather that such variables do not tell the whole story, so that "true" productivity has other components as well. One need only think of examples in which the employer claims to observe on the job work effort or "attitude" to see the point.

But if (9) holds, then, substituting in (5), we obtain

$$Y = G\alpha^* + X(H\beta^*) + (e^{**} + W\beta^*), \tag{10}$$

and this is in the same form as (1) with the presence of $W\beta^*$ in the error term causing no difficulty, because it is uncorrelated with X. In this case (which I strongly believe to be the likely one), direct regression leads to a consistent estimate of α^*, whereas reverse regression certainly does not. Indeed (under the same conditions as before), reverse regression is likely to lead to an estimate of α^* that is biased downward, both because of the presence of e^{**} on the righthand side of (8) as before *and* because of a similar effect from the fact that $W\beta^*$ will now also appear in the error term of that equation.

In passing, note that, as (9) implies, further information possessed by the employer will be relevant *only* to the extent that it is orthogonal to X. For example, if the employer claims to observe on the job performance or "attitude," and years of education are included in X, merely knowing that men and women score differently on the employer's measure does not invalidate the use of (1). Because education may affect such scores, the employer's information only adds something if male-female score differentials persist when education effects are held constant. Even then, as just shown, estimation of (1) remains consistent.

Reverse regression, then, is unlikely to be an attractive way to deal with omitted variables in the present context. Its proponents, however (including Dempster), have put forth another justification for its use. This is the possibility that the firm engages in what Conway and Roberts (1984, page 128) call "Hiring 2," choosing wage (and other job characteristics) first and then hiring the job applicant with the highest qualifications. Such a procedure leads directly to reverse regression as a test for "Fairness 2"—the carrying out of this procedure in a sex-blind way.

There are several things to be said about this argument. First, if the firm really does engage in "Hiring 2," one of the arguments given above for the biased nature of reverse regression will certainly fail. If qualifications are conditioned on wages and not the other way round, the the appearance of e^{**} in (8) causes no difficulty. Indeed, with "Hiring 2," a parallel argument shows that direct regression leads to biased results.

This does not rescue reverse regression, however, for the considerations as to the nature of omitted variables remain the same whether or not "Hiring 2" is involved. To see this, return to the univariate case. "Hiring 2" means the employer chooses X^* by

$$X^* = G\lambda + Y\mu + u, \tag{11}$$

where u is a random disturbance and λ and μ are parameters. When the statistician uses X in place of X^*, then (taking $H = 1$ in (9) for simplicity)

$$X = G\lambda + Y\mu + (u - W). \tag{12}$$

But W will certainly be negatively correlated with Y, and, as before, this will lead to underestimating the extent of discrimination.

Returning to the multivariate case, this shows that reverse regression (as Dempster recognizes) will not be an adequate test for "Fairness 2," unless $X^*\beta^*$ rather than X is used. This, however, brings us back to the question of what forms of argument should be acceptable. To justify reverse regression (and, indeed, to combat the results of direct regression), the employer must be able to specify X^*. Further, for present purposes, he must specify β^*. Even if he does that, however, one ought not to accept reverse regression as the appropriate method without a credible showing by the employer that he practices "Hiring 2"—choosing qualifications given wages—rather than "Hiring 1"—paying wages given qualifications. If "Hiring 2" really is used, such a showing ought not to be burdensome. Large firms tend to have written personnel policies. Even small ones can give examples of their job advertising and anecdotal testimony from personnel officers. Lacking such a showing, one ought to suspect that reverse regression—as its intellectual history suggests—is just an answer looking for a good question.

This is related to the final question I shall discuss, that of what Dempster calls "judgmental discrimination." Put aside all issues of statistical method, says Dempster. Suppose that "a presumed honest attempt to assess productivity" involves a positive α'' in (3). Then even knowledge of α^* in (5) will not suffice as a test for sex discrimination, because a positive α^* may simply reflect a positive α''.

To accept this argument is to accept anything as a defense and always find for the defendant. A positive α'' means that the employer conditions wages on gender, *given* all the other information that he has. It means that women are seen to be less productive, not because they differ from men in education or measurable skills, or even because they want different hours or conditions of travel for child-rearing considerations. Any of those propositions could be tested because neither women nor men are all alike in such dimensions. (For example, some men are single parents and some women are not married.) Of course it is *possible* that the employer "knows" that women and men differ for nontestable reasons, but (as in all cases of untestable private information) that ought not to be an acceptable form of argument.

Let me put this in plain English. Acceptance of "judgmental discrimination" as a legitimate defense has nothing to do with statistics. It means allowing the employer to defend by saying: "Of course I'm not prejudiced against women (blacks). It's just that—*for reasons I can't explain*—they can't do a (white) man's job." Experts ought to think hard before lending themselves to positions like this.

ACKNOWLEDGMENTS

My original testimony on this subject benefited from conversations with Richard Gilman, Diane Heim,

Bruce Levin, Debra Millenson and Mary Ellen Wynn. I retain responsibility for error.

ADDITIONAL REFERENCES

CONWAY, D. A. and ROBERTS, H. V. (1984). Rejoinder to comments on "Reverse regression, fairness and employment discrimination." *J. Bus. Econ. Statist.* **2** 126–139.

FERBER, M. A. and GREEN, C. A. (1984). What kind of fairness is fair? A comment on Conway and Roberts. *J. Bus. Econ. Statist.* **2** 111–113.

ROBERTS, H. V. (1979). Harris Trust and Savings Bank: An analysis of employee compensation. Report 7946, CMSBE, Graduate School of Business, Univ. Chicago.

ROBERTS, H. V. (1980). Statistical biases in the measurement of employment bias. In *Comparable Worth: Issues and Alternatives* (E. R. Livernash, ed.). Equal Employment Advisory Council, Washington.

Comment

Arthur S. Goldberger

I am grateful to Arthur Dempster for pointing out an error in my article, but perturbed by his campaign against econometricians. On balance, my perturbation exceeds my gratitude.

A bit of background. The most popular approach to the assessment of gender discrimination has been to run the direct regression

$$\hat{y} = \mathbf{b}'\mathbf{x} + az,$$

where y = salary, $\mathbf{x}$ = vector of measured covariates and z = gender (coded 1 for men, 0 for women). In this approach, the coefficient "a" (typically positive) is taken to be the measure of discriminatory behavior on the part of the employer. An obvious objection is that relevant covariates have been omitted: $\mathbf{x}$ may not capture all the productivity-related characteristics available to the employer. When those covariates are correlated with gender, there is a presumption that their omission biases the direct regression estimate.

Some critics of direct regression had gone on to suggest that the bias would be eliminated by using reverse regression, in particular by running the composite covariate $q = \mathbf{b}'\mathbf{x}$ upon y and z,

$$\hat{q} = cy + dz,$$

and taking $-d/c$ as the measure of discriminatory behavior. The rationale for this was rather vague, some mention of errors in variables being made.

To an econometrician, it seemed inappropriate to discuss estimation bias until the parameter of interest had been defined and imbedded in a coherent model. I first sought a model that would support direct regression, and yet allow for omitted variables in the statisticians's regression. I found it in Model A:

$$\text{(A1)}\qquad y = p + \alpha z,\quad p = \mathbf{x}'\boldsymbol{\beta} + w,\quad \mathbf{x} = \boldsymbol{\mu} z + \mathbf{u},$$

$$\text{(A2)}\qquad E(w \mid \mathbf{x}, z) = 0,\quad E(\mathbf{u} \mid z) = \mathbf{0}.$$

The parameter of interest is α. I wrote that p is the "latent variable which is best interpreted as the employer's assessment of productivity" and that w is "a gender-free disturbance. That disturbance represents the additional information available to the employer but not to the statistician."

In this model, I deduced that

$$\text{(A3)}\qquad E(y \mid \mathbf{x}, z) = \mathbf{x}'\boldsymbol{\beta} + \alpha z,$$

so that "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." The key to the conclusion is the assumption that $E(w \mid \mathbf{x}, z) = 0$—the omitted variables are uncorrelated with gender after controlling for the measured variables. That is precisely why I introduced it.

Next I sought a model that would support reverse regression. Drawing on suggestions made by proponents of reverse regression, I found it in Model B:

$$\text{(B1)}\qquad y = p + \alpha z,\quad \mathbf{x} = \boldsymbol{\gamma} p + \boldsymbol{\varepsilon},\quad p = \mu z + u,$$

$$\text{(B2)}\qquad E(\boldsymbol{\varepsilon} \mid p, z) = \mathbf{0},\quad E(u \mid z) = 0.$$

I wrote that here "each observed qualification [element of $\mathbf{x}$] is merely an indicator of the employer's assessment [p] subject to a gender-free disturbance." The parameter of interest is again α.

Arthur S. Goldberger is Vilas Research Professor of Economics, Department of Economics, University of Wisconsin, Madison, Wisconsin 53706.