Comment

Delores A. Conway

Dempster has written several papers on the statistical assessment of employment discrimination that offer key insights and help to identify critical problems in the area. For example, a number of years ago, I was struck by a quotation from Judge Bazelon, that appeared in an early working paper (Dempster, 1979) and has served as a "guiding light" for my own research. Commenting on the role of scientists in legal cases, Bazelon (1979) asserts, "In the scientists' realm—the sphere of fact—courts can ask that the data be described, hypotheses articulated, and above all, in those areas where we lack knowledge, that ignorance be confessed."

A unifying and important theme in Dempster's papers concerns how incomplete knowledge of the employment process may seriously bias conclusions about possible discrimination. The current article continues the tradition and offers a probing analysis of data limitations and the impact that different assumptions have on conclusions. He further clarifies ideas from his previous papers by proposing a general model of the employment process.

The first two sections of the article highlight the need to understand causal mechanisms underlying statistical models in observational studies. The distinction between the use of chance mechanisms as "analogies" and their use as "realities" is an important one. Too infrequently in observational studies do we think carefully about the random mechanism generating the data. As a result, there is a great need to develop formal frameworks that combine the information in the observed data with critical background information. I share Dempster's optimism that we are making progress in this area—evidenced by recent developments in statistical graphics, computer-intensive methods and implementations of Bayesian analysis. However, the development of suitable frameworks for analyzing observational data will remain a major challenge for the profession in many years to come.

A separate section of the article addresses the issue of "judgmental discrimination." This is a controversial subject that requires considerable thought and reflection. It also involves important legal issues that draw a distinction between disparate impact and treatment (see, e.g., Brodin, 1982; Furnish, 1982; Lopatka, 1977; Manishin, 1980). My major comments target the assumptions and implications of the general model of the employment process presented in the third section. Also, I will focus attention on estimating the total discrimination effect and bypass the issue of "judgmental discrimination."

1. ILLUSTRATION OF CAUSAL CONCEPTS WITH A DATA APPLICATION

The general model of the employment process proposed by Dempster appears in equation (9) of the article. The model assumes that the observed qualifications $\tilde{X}_1$ are independent of the unobserved qualifications $\tilde{X}_2$ and that α^* measures the total discrimination effect. Furthermore, both direct and reverse regression provide assessments of possible sex discrimination, given by α and α_R, respectively.

I find it easier to understand the assumptions and implications of Dempster's causal model in the context of an actual data application. The first data set consists of 274 white, professional employees hired at a large bank between 1971 and 1972. The data are a small component of a much larger study that was developed for a legal case involving possible salary discrimination against females. Regression analyses from this data appear in Conway and Roberts (1983).

The natural log of 1976 salary was regressed on four available job qualifications, linearizing transformations of the basic qualifications and sex. (Consistent with Dempster's notation, sex is an indicator variable that equals 1 for males and 0 for females.) The variables ED7, ED8 and ED9 are categorical variables for educational levels. WORK is the 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 unconditional mean salary difference from the data is $\bar{Y}_M - \bar{Y}_F = 0.202$, suggesting that the average salary is about 20% higher for males than females. The observed job qualifications help to account for part of this difference, because the mean qualification difference from the direct regression of log salary on the above qualifications is $\bar{X}_{1M} - \bar{X}_{1F} = 0.054$. This suggests that the average qualification index is 5.4% higher for males than females.

The direct regression estimate of α is 0.148 with a standard error of 0.0356. This suggests that females have an estimated salary shortfall of 14.8% and the result is statistically significant. The estimated sex and salary coefficients from reverse regression are

Delores A. Conway is Associate Professor of Statistics at the School of Business Administration, University of Southern California, Los Angeles, California 90089.

-0.0097 and 0.316, respectively, so α_R is estimated by $0.0097/0.316 = 0.031$. The results from reverse regression suggest that females have an estimated salary shortfall of 3.1%, although the result is not statistically significant. The two different estimates are typical of those found in similar studies, where direct regression often suggests a substantial shortfall in female salaries adjusted for qualifications and reverse regression indicates approximate parity.

The question arises as to which estimate $\alpha = 0.148$ or $\alpha_R = 0.031$ measures the total discrimination effect α^*. In actual legal cases, differences in the assessment of possible discrimination of this order of magnitude are substantial and can have large monetary implications for back pay settlements. At this point, underlying causal models enter the picture to help evaluate which estimate more accurately approximates α^*.

Dempster's model implies that the choice among α, α_R or something completely different depends upon external assumptions about the unobserved mean qualification difference $\mu_{M2} - \mu_{F2}$. Equation (12) states that the direct regression assessment α is unbiased when $\mu_{M2} - \mu_{F2} = 0$. Similarly, (15) and (16) show that the reverse regression assessment α_R is unbiased when $\mu_{M2} - \mu_{F2} = (\tau_2/\tau_1)(\mu_{M1} - \mu_{F1})$, where $\mu_{M1} - \mu_{F1}$ is the observed mean qualification difference.

We can estimate the size of the required difference in (16) from the observed qualifications. A simple derivation shows that the righthand side of (16) equals

$$(\tau_2/\tau_1)(\mu_{M1} - \mu_{F1}) = [(1 - \beta^*)/\beta^*](\mu_{M1} - \mu_{F1}) = [(1 - R^2)/R^2](\mu_{M1} - \mu_{F1}),$$

where β^* is the slope coefficient in the reverse regression of $\tilde{X}_1$ on Y and $R^2 = \mathrm{Corr}^2(X_1, Y \mid \mathrm{Sex})$ is the within-groups multiple correlation coefficient between $\tilde{X}_1$ and Y (see, e.g., Conway and Roberts, 1984, page 130). An even simpler form is given by $(\tau_2/\tau_1)(\mu_{M1} - \mu_{F1}) = \alpha - \alpha_R$, which follows from (12) and (15).

The implications for the group of 274 professional employees are interesting to consider. The direct regression assessment of a 14.8% female salary shortfall accurately reflects the total discrimination effect if $\mu_{M2} - \mu_{F2}$ is zero. By contrast, the reverse regression assessment of essential parity in salaries is accurate when this difference is approximately $14.8 - 3.1 = 11.7\%$. This means that the unobserved characteristics have an additional sex effect that is approximately twice as large as the observed qualification difference of 5.4%.

It is well known that the direct and reverse regression assessments of possible discrimination can change with the inclusion of omitted variables. We can demonstrate such changes for the bank employees by introducing the observed qualifications one at a time into the regression model. The first two columns of Table 1 give the direct and reverse regression assessments of fairness for a hierarchy of models in which each qualification is added to the previous set in succession. For example, the first model in Table 1 uses only ED7 as the set of observed qualifications; the second uses ED7 and ED8; the third uses ED7, ED8, and ED9 and so forth.

The results show that the direct regression sex coefficient declines from a female salary shortfall of 21.4% in the first model to a shortfall of 14.8% in the last model. The comparable reverse regression assessment cannot be distinguished from zero in all the models. The strength of association between Y and $\tilde{X}_1$ is given by the reverse regression slope coefficient b^*, which equals the sample within-groups multiple correlation coefficient. As qualifications are added to the model, the multiple correlation coefficient increases from 8.7% to 31.6%. Furthermore, the observed qualification difference between males and females also increases from approximate parity in the first model to 5.4% in the final model.

The last two columns of Table 1 estimate the condition in Dempster's model for reverse regression to

TABLE 1

Direct and reverse regression assessments of fairness for an entering cohort of 274 professionals from a large bank when job qualification variables are introduced successively

Variable included	Estimated sex differential			Qualification difference multiple		
	a (direct)	a_R (reverse)	b^*	$\bar{X}_{1M} - \bar{X}_{1F}$	$(1 - b^*)/b^*$	$a - a_R$
ED7	0.203[a]	0.212	0.087	−0.001	10.44	−0.010
ED8	0.191[a]	0.108	0.123	0.012	7.14	0.086
ED9	0.159[a]	−0.006	0.208	0.043	3.81	0.164
SENSQ	0.165[a]	0.044	0.237	0.038	3.21	0.122
WORK	0.156[a]	0.021	0.257	0.047	2.88	0.136
WK/AGE	0.148[a]	0.031	0.316	0.054	2.16	0.117

$\bar{Y}_M - \bar{Y}_F = 9.946 - 9.744 = 0.202$

[a] Statistically significant at a 0.01 level.

provide an unbiased estimate of total discrimination. The accuracy of the first two models is questionable, because so little of the variation in log salaries is related to the two indicator variables. For the third through final models, the unobserved qualification difference $\mu_{M2} - \mu_{F2}$ would be close to 12% if the reverse regression assessment of approximate parity in adjusted salaries is correct. Notice that the multiple of the observed qualification difference decreases steadily from 3.8 to 2.2. This reflects the fact that R^2 is steadily increasing, so that $(1 - R^2)/R^2$ declines. The multiple is exactly 1, when $R^2 = 0.5$ and decreases to zero as R^2 approaches 1.

2. THE EFFECTS OF OMITTED CHARACTERISTICS

The results from a second set of data used in a legal case are very similar. The second data set consists of 275 clerical and 80 professional employees hired at Harris Bank between 1969 and 1971. Extensive analysis of this data appears in Conway and Roberts (1986a). Again possible salary discrimination against females is the question of interest.

The set of observed qualifications consists of four basic proxies: education, age, work experience and seniority. The Y variable refers to log salary in 1977. The unconditional salary difference $\bar{Y}_M - \bar{Y}_F$ is 24.7 and 31.2% for the clerical and professional groups, respectively.

Table 2 gives the results for the two groups when observed qualifications are added successively to a regression model, in the same way as for Table 1. The results are very similar to those in Table 1. Both the direct and reverse regression assessments of possible discrimination conflict, with direct regression reflecting an estimated female salary shortfall of 14.8 and 12.1% for clerical and professional employees, respectively, in the final models. The reverse regression assessment shows approximate parity for the clerical group and a male shortfall of 35.3% in salaries for the professional group. The result for the professional group is somewhat suspect, because there are only 10 females in this group.

The strength of the relationship between Y and $\tilde{X}_1$ increases within the two job groups, as does the observed qualification difference $\bar{X}_{1M} - \bar{X}_{1F}$. The last two columns of Table 2 indicate that the unobserved characteristics would have fairly substantial sex effects, if the reverse regression assessment accurately reflects the total discrimination effect. For example, the estimate of 0.47 required for $\mu_{M2} - \mu_{F2}$ in the final model of the professional group is quite large.

The results for the two data sets are interesting and help to clarify some causal implications from Dempster's model. First, it is important to recognize that $a - a_R$ is only a point estimate of the righthand side of (16). The standard error of this estimator is undoubtedly substantial. Time did not permit my computation of the standard errors, but they should be part of the analysis. In fact, we could approximate the finite sampling distribution of $a - a_R$ by a bootstrap procedure and obtain a much better idea of the likely values.

Second, computation of the observed qualification difference $\bar{X}_{1M} - \bar{X}_{1F}$, the multiple $(1 - b^*)/b^*$ and the difference in assessments $a - a_R$ brings out some new perspectives of the analysis. Many employment

TABLE 2
Direct and reverse regression assessments of fairness for clerical and professional employees hired at Harris Bank in 1969–1971 when job qualification variables are introduced successively

Variable included	Estimated sex differential			Qualification difference multiple		
	a (direct)	a_R (reverse)	b^*	$\bar{X}_{1M} - \bar{X}_{1F}$	$(1 - b^*)/b^*$	$a - a_R$
a. 275 clerical employees						
Some college	0.214[a]	−1.160	0.023	0.033	41.58	1.374
College degree	0.204[a]	0.040	0.208	0.043	3.82	0.164
Graduate degree	0.194[a]	0.032	0.247	0.053	3.05	0.162
Age	0.179[a]	0.044	0.336	0.068	1.97	0.135
Work experience	0.181[a]	0.052	0.337	0.066	1.97	0.130
Work Sq.	0.159[a]	0.001	0.359	0.088	1.78	0.157
Seniority	0.148[a]	−0.006	0.391	0.099	1.56	0.154
b. 80 professional employees						
College degree	0.304[a]	−0.718	0.007	0.007	137.12	1.023
Master's degree	0.297[a]	−0.835	0.013	0.015	76.11	1.131
MBA degree	0.167[a]	−0.386	0.207	0.144	3.84	0.554
Age	0.158[a]	−0.394	0.217	0.153	3.60	0.552
Work experience	0.153[a]	−0.274	0.271	0.159	2.69	0.427
Work Sq.	0.135[a]	−0.323	0.279	0.177	2.58	0.458
Seniority	0.121[a]	−0.353[a]	0.286	0.190	2.49	0.474

[a] Statistically significant at a 0.01 level.

studies record values of R^2 in the range between 20 and 50% (see, e.g., Borjas, 1978; Triemann and Hartmann, 1981). Dempster's model then implies that the reverse regression assessment of total discrimination will be unbiased if the sex differences in the unobserved characteristics exceed those of the observed qualifications. Consequently, we would expect that $a - a_R$ exceeds $\bar{X}_{1M} - \bar{X}_{1F}$ in many studies.

Direct regression proponents might conclude that this result is implausible and casts doubt on the reverse regression assessment as an accurate measure of the total discrimination effect. On the other hand, reverse regression proponents might argue that this only demonstrates the magnitude and impact of the omitted factors not taken into account by the regression model. The courts have generally ruled that assumptions about unobserved variables must be viable (i.e., there is a real effect omitted). Furthermore, exclusion must systematically bias conclusions about fairness based on the observed data (see, e.g., Vuyanich versus Republic National Bank, 1980).

Dempster's general model helps focus attention on the magnitude of the bias due to omitted variables. It helps the analyst consider what kind of omitted variables might have sex effects of a specified magnitude that is pertinent to assessments of fairness. For example, could market factors or level of responsibility account for an unobserved sex differential of 20% for the professional employees in Table 1? Some recent work shows that failure to account for differences among job positions can lead to substantial biases in regression assessments. This suggests an analysis based on disaggregated groups that are relatively homogeneous (see, e.g., Conway and Roberts, 1987).

Third, Dempster's general model might be used to modify the two regression assessments α and α_R. Reliable background information about the unobserved variables may specify a range of values for $\mu_{M2} - \mu_{F2}$ that differs from either zero or $\alpha - \alpha_R$. This would suggest that the model could be used to obtain a more plausible range for the total discrimination effect that differs from either α or α_R, but is instead a weighted average of the two. The exact weights on the different regression assessments could be derived from (12) and (15).

3. GOLDBERGER'S MODELS OF THE EMPLOYMENT PROCESS

Dempster considers the relationship between the model of the employment process in (9) and Goldberger's Models A and B. The comments that delineate the relationship between his model and Goldberger's Model A are completely correct. I also agree that the proportionality constraint implied by Model B follows directly from equation (40b). The introduction of sex effects by a system of equations of the form,

$$x = \theta \,\mathrm{Sex} + \gamma p + \varepsilon,$$

does eliminate the proportionality constraint. Furthermore, the condition that the reverse regression assessment is unbiased for the above system cannot be tested empirically with the data. (The derivation is straightforward and is omitted.)

However, I disagree with Dempster that the generalized Model B given above is "stochastically equivalent to my general model." The difference arises from the assumption that $\tilde{X}_1$ and $\tilde{X}_2$ in (9) are uncorrelated. It is clear from (40a) and (40b) that the omitted factors $\tilde{X}_2$ in Goldberger's Model B are correlated with $\tilde{X}_1$ (and uncorrelated with Y). The correlation between the observed characteristics $\tilde{X}_1$ and the random disturbance in the two models is quite different which results in different implications about the bias in α and α_R. Consequently, I do not believe that the two models are equivalent, but instead imply a different random mechanism generating the data.

In fact, Dempster's general model seems most closely related to a process that Harry Roberts and I have called "Type-1 Employer Behavior" (Conway and Roberts, 1986a). This is due to the fact that "the linear compounds represented by $\tilde{X}_2$ are uncorrelated with X," and hence $\tilde{X}_1 = X\beta'$. This assumption is also a critical one in Goldberger's Model A. However, Model B characterizes a process that we have called "Type-2 Employer Behavior," where the observed qualifications X are related to the unobserved characteristics $\tilde{X}_2$.

4. ALTERNATIVE METHODS OF ANALYSIS FOR EMPLOYMENT STUDIES

One of the advantages of the direct and reverse regression approaches, used in tandem, is that it enlarges our perspective of the employment process. Conflicts in the two assessments of possible discrimination often highlight the limitations of data and our own knowledge about this process. To develop more accurate assessments of employment discrimination, I feel that we need additional sources of data and methods of analysis to supplement the causal models and regression methods of the last 10 years.

For example, certain legal cases have resulted in large back pay settlements to remove salary differentials that were judged to be discriminatory. It would be interesting to analyze the relationship between salaries and observed qualifications both before and after the intervention. This might help to shed additional light on the bias in conclusions from different regression assessments.

Recently, a number of large companies have expanded the amount of information on employee qualifications collected for salary, hiring and promotion decisions. In some companies, the decisions also involve multiple raters (which might include a self-evaluation), so that there are multiple measures of assessed productivity. Assessment of possible employment discrimination would undoubtedly be easier with a richer data base that included unbiased assessments of productivity and would suggest new methods of analysis.

Finally, I know of at least one company that uses an explicit direct regression approach for assigning salaries by a computer algorithm to insure fairness 1. It would be interesting to examine the data from such companies over time to help isolate market factors that might affect different job positions. It would also be of interest to evaluate the implications of this approach with respect to fairness, personnel costs, quality of the workforce and market competition.

As more sophisticated information becomes available for employment studies, our methods of analysis should also expand to exploit this information in creative ways. In this way, we will continue to improve our understanding of the employment process and foster the development of more realistic models.

ADDITIONAL REFERENCES

BAZELON, D. L. (1979). Risk and responsibility. *Science* **205** 277–280.

BORJAS, G. J. (1978). Discrimination in HEW: Is the doctor sick or are the patients healthy? *J. Law Economics* **21** 97–110.

BRODIN, M. S. (1982). The standard of causation in the mixed-motive Title VII action: A social policy perspective. *Columbia Law Rev.* **82** 292–326.

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.

CONWAY, D. A. and ROBERTS, H. V. (1986a). Regression analyses in employment discrimination cases (with discussion). In *Statistics and the Law* (M. H. DeGroot, S. E. Fienberg and J. B. Kadane, eds.) 107–195. Wiley, New York.

CONWAY, D. A. and ROBERTS, H. V. (1987). Analysis of employment discrimination through homogeneous job groups. Working paper, School of Business, Univ. Southern California.

DEMPSTER, A. P. (1979). Statistical concepts of discrimination. Working paper, Dept. Statistics, Harvard Univ.

FURNISH, H. A. (1982). A path through the maze: Disparate impact and disparate treatment under Title VII of the Civil Rights Act of 1964 after Beazer and Burdine. *Boston College Law Rev.* **23** 419–445.

LOPATKA, K. T. (1977). A 1977 primer on the federal regulation of employment discrimination. *Univ. Illinois Law Forum* **1977** 69–168.

MANISHIN, G. B. (1980). Section 1981: Discriminatory purpose or disproportionate impact. *Columbia Law Rev.* **80** 137–170.

TRIEMANN, D. J. and HARTMANN, H. I., eds. (1981). *Women, Work and Wages: Equal Pay for Jobs of Equal Value.* National Academy Press, Washington.

VUYANICH VS. REPUBLIC NATIONAL BANK (1980). F.E.P. Cases 128 (N.D. Texas).

Comment

Joseph L. Gastwirth

Professor Dempster's interesting and thought-provoking article concerns many scientifically fundamental and socially important issues: the relationship between statistical inference and causality, the proper construction of probabilistic models, the effect of omitted predictor variables (OVs) and errors in variables (EVs) on inferences drawn from models fitted to data and the implications of these topics on the statistical analyses relied on in employment discrimination cases and related public policy decisions. His formulas (12) and (15) concerning the effect of omitted variables on direct and reverse regressions, respectively, add to our understanding of these techniques. From my analysis (Gastwirth, 1984, 1988) of EEO cases, I believe Prof. Dempster may have overestimated the potential for "legal mischief" although I agree with him that a proper statistical analysis involves a careful evaluation of the data and model, including consideration of data errors and omitted factors. Indeed, the effects of OVs stemming from Cornfield's analysis of the possible effect of OVs on the smoking and lung cancer association (see Greenhouse, 1982, for details; Rosenbaum, 1987, for recent developments) and the Bayesian view of missing evidence (Lindley and Eggleston, 1983) have implicitly been used by policy makers and judges. I wish Dempster's use of the Bayesian approach was more explicit so we could compare his conclusions with those reached by the judiciary in actual cases. In particular, the process used by the employer in computing the "posterior expected reward . . . employee" is precisely what is at issue in a disparate treatment case.

Dempster emphasizes the importance of careful causal modeling and considers the randomized clinical trial the most convincing statistical design. However, he also notes that the decisions examined

Joseph L. Gastwirth is Professor of Statistics and Economics, Department of Statistics, George Washington University, Washington, D. C. 20052.