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."

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

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