

Comment

Harry V. Roberts

We appear to have passed through the period of most intense litigation of legal cases alleging employment discrimination in compensation or promotion by race or sex. The highest frequency of academic papers on the statistics of age discrimination appears also to be past.

Arthur Dempster's paper therefore comes at a good time for an overview of the statistical issues raised by race and sex cases. In my view, Dempster's overview is fascinating, authoritative and close to comprehensive on the foundations of discrimination study with statistical tools. The paper not only deals with discrimination studies in particular, but illuminates the vexing general problem of causal inference from observational data where explicit randomization has not or even cannot be applied.

The paper warns also about the dangers in the belief that sufficiently sophisticated models can completely resolve differences over causal interpretation: "statistical analysis on its own rarely offers complete solutions to externally specified problems." Dempster stresses the need to "look outside the data for evidence bearing on the missing information."

I find nothing to criticize in Dempster's general development or in his model based on Bayesian principles. At the same time, I feel that more attention to certain concrete details would help to deal more specifically with important questions that have been debated by statisticians inside courtrooms and in the pages of scholarly publications. I shall therefore attempt to fill in details that I believe to be important.

Part of my comment will be directed to an amplification of Dempster's comment about "looking outside the data." Looking outside the data entails examination of the employment process that produces the data. Such examination reveals specific activities such as search for job candidates, selection from applicant pools, initial placement and salary determination, salary advancement and job promotion and even reduction of force, which has come to be important in the statistical analysis of age discrimination cases. In looking at these activities, one is led to flesh out the econometric model of Section 3. One is led also to consider legal questions. One is made more aware of the limitations of available data, including both job

qualifications and employment history. For example, good job performance measures are seldom available and information is often lacking about job applicants who were not hired.

But one can also look *within the data*: mundane data analysis can provide clues about causal analysis. It was through data analysis in a series of discrimination cases from 1974–1985 that I gradually acquired a better understanding of what happens in the employment process that is relevant to a statistical study of the process. That understanding led to modifications and extensions of the methodology of earlier studies, including first the introduction of reverse regression and later the modification of the way reverse regression is applied. Consideration of problems that I encountered in data analysis leads to a fuller understanding of the issues so well presented by Dempster. (This approach also gives an autobiographical tone to my comment that I regret but cannot avoid without blunting the points to be made. I regret also the necessarily extensive self-citations in the references for this comment.)

During the process of data analysis, not only my statistical methodology but my views of the important causal issues evolved. The evolution is reflected in the series of papers and reports listed under "Additional References." (The papers from 1983 on were jointly authored with Delores A. Conway. I cannot overstate the importance of Professor Conway's contributions to my own thinking, but because we are preparing separate comments on Dempster's paper that we have not had an opportunity to discuss with each other, I hasten to say that she may not fully agree with all aspects of this comment.)

ADVANCEMENT STUDIES

In 1974–1975 I worked with Mary Townsend Kimpton on a salary discrimination case involving United Airlines (Field, Kimpton, McGee and Roberts, 1978). Study of salary discrimination at hire was constrained by lack of relevant information on job qualifications. Hence, salary advancement after hire seemed to be the only aspect of the case amenable to regression analysis. Hence, our regression study conditioned on starting salaries and on the available (limited) information about job qualifications. I came later to refer to such a study as an "advancement study."

It is interesting to observe that at that time, the plaintiff presented only simple tabular comparisons of

Harry V. Roberts is Professor of Statistics, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637.