

discrimination on employee qualifications, but that is a different problem.

IS GENDER A CAUSE?

I have argued elsewhere (Holland, 1986a) that gender is not usefully thought of as a cause in many social science applications, and I would like to point out that I (and, I believe, Dempster) have remained true to this position in the present discussion. The “causes” involved here are discriminatory practices in salary administration, not the genders of the people involved. It is true that gender plays a role in the causal theory (B), but only in the sense that the causal effect of discrimination varies with the gender of the employee (which is, after all, what *discrimination* means). This distinction is blurred in the regression function, $E(Y_d | G, X_d) = k + \alpha G + X_d \beta$, where one is apt to call α the “effect” of G on Y_d . This is unfortunate usage and is often a source of confusion in the casual causal talk that often accompanies regression analyses. Dempster is to be admired for avoiding such a casual approach to causation.

CONCLUSIONS

I hope I have sketched enough to show that the use of Rubin’s model, with its focus on the measurement of causal effects, can be used to produce a crisp analysis of the employment discrimination problem that is very similar to much of that given by Dempster but without his need to interpret Y_c as the result of an optimal decision rule used by a thoughtful employer who invokes posterior means, loss functions and prior distributions. $Y_c(u)$ is a crucial number that we usu-

ally do not observe and which, because of this, can easily be swept under the rug and forgotten. Who really knows how Y_c should be determined? Is it possible to make serious efforts to actually *measure* some Y_c values rather than to continue to make them up? Perhaps there are some firms or parts of firms that do not discriminate in their administration of salaries; could their data be used to study Y_c directly in some specialized situations? On the other hand, because of the difficulty (and, often, the impossibility) of measuring Y_c , it should be clear that the analysis of employment discrimination differs significantly from the standard observational study in which the responses of *both* treated and control cases are always obtained. A regression analysis done either forward or backward cannot solve this fundamental problem with the analysis of employment discrimination.

I believe that the problem of employment discrimination is both serious and complex. It surely deserves a better effort than a parade of tired, old regression “paradoxes” by well intentioned men and women through countless courtrooms; if such a parade is the best that statistical science can do, perhaps it is doing more harm than good.

ADDITIONAL REFERENCES

- HOLLAND, P. W. (1986b). Which comes first, cause or effect? *The New York Statistician* **38** 1–6.
- HOLLAND, P. W. and RUBIN, D. B. (1983). On Lord’s paradox. In *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord* (H. Wainer and S. Messick, eds.). Lawrence Erlbaum Associates, Hillsdale, N. J.
- RUBIN, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Ann. Statist.* **6** 34–58.

Comment: Statistical Science and Economic Science

John Geweke

Professor Dempster has argued in favor of constructing models that explicitly specify stochastic components, and against the alternative of using models that introduce convenient but *ad hoc* chance

mechanisms. There is increasing recognition among academic econometricians that this explicit specification is necessary for a model to be causal, that is, for a model to evaluate counterfactuals reliably and therefore to be employed for the purpose of policy evaluation. Explicitly specified stochastic components often arise from economic agents having information sets broader than analysts’ information sets, as in Dempster’s approach. A very successful application of this strategy is the development of asset pricing

John Geweke is William R. Kenan, Jr., Professor of Economics and Professor of Statistics and Decision Sciences, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27706.