latent cause is often treated as a treatment measured with error. Right now I don't have an answer as to whether or not latent variables can be causes. I suspect that it might depend on other considerations. But I am sure that it is at least as important to know the answer to this question as it is to know how to fit a very complicated structural model by maximum likelihood.

Why go on and on about causal models and latent variables? Here is my answer. If, as Clogg rightly asserts, sociological research *is* influencing statistical research in the study of causation, what, then, should this influence be? Should statisticians jump on the band wagon and develop more and more procedures for fitting these models, following the path so well blazed by, say, Anderson and Rubin (1956)? Should the reaction of statistical research be simply to continue to add to the list of structural models that can be fit to data in finite computer time? Or should it spend some effort to give these models a better foundation based on the known past successes of statistical science? In my view, the latter is one of the many contributions of the approach to causal inference that Rubin started. His approach grows out of work in many fields all bearing on the problem of causal inference when there is heterogeneity, variation and noise – *Unruly's feast* – and there is plenty more to do there. Again, it's the interplay between statistics, science and mathematics that's important.

ACKNOWLEDGMENTS

These comments were prepared while the author was a 1991–1992 Fellow at The Center for Advanced Study in the Behavioral Sciences and supported, in part, by the Spencer Foundation.

Comment

Charles F. Manski

I can easily understand Professor's Clogg's frustration with the belief that advances in statistics "trickle down" to the social sciences. Statisticians must feel the same way when it is said that advances in probability theory trickle down to statistics. Clogg's account of the historically productive two-way flow of ideas between statisticians and sociological methodologists is well written and instructive. It is easy enough to document a similarly productive flow of ideas between statisticians and econometricians. Some examples follow.

COLLABORATION BETWEEN STATISTICIANS AND ECONOMETRICIANS

In the 1940s and 1950s, statisticians and econometricians concerned with the estimation of linear model systems worked closely together with the support of the Cowles Commission for Research in Economics. The statisticians in the group included, among others, Ted Anderson, Herman Chernoff, M. A. Girshick and Herman Rubin. The economists included, among others, four later winners of the Nobel Prize: Trygve Haavelmo, Lawrence Klein, Tjalling Koopmans and Herbert Simon. The Cowles Commission work revolutionized econometrics. See the seminal volume edited by Hood and Koopmans (1953). The atmosphere and substance of the collaboration between statisticians and econometricians is conveyed well by Anderson (1991), written on the occasion of Haavelmo's receipt of the Nobel Prize.

From the 1970s through the present, Bayesian statisticians and econometricians have met on a regular basis. An important medium for these contacts has been the conference series organized by Arnold Zellner, a prominent econometrician who is currently the president of the American Statistical Association. Several published volumes have emerged from these conferences. See, for example, Fienberg and Zellner (1975).

Over the past 10 years, statisticians and econometricians working in the area of nonparametric and semiparametric analysis have developed increasingly close working relationships. The fruits of collaborative research have appeared in co-authored articles, such as Heckman and Singer (1984) and Pakes and Pollard (1989). Several conferences have brought together statisticians and econometricians, with tangible product in the form of conference volumes such as that edited by Barnett, Powell and Tauchen (1991). Knowledge of mutually interesting developments has also diffused through the routine process of exchanging working papers.

Charles F. Manski is the Wolfowitz Professor of Economics, University of Wisconsin–Madison, 1180 Observatory Drive, Madison, Wisconsin 53706.

My own research and teaching during the past 10 years have been heavily influenced by the work of statisticians on nonparametric regression (e.g., Stone, 1977), semiparametric efficiency analysis (e.g., Bickel, 1982; Begun et al., 1983) and empirical process theory (e.g., Pollard, 1984). Statisticians have in turn made use of the contributions of econometricians on such topics as quantile regression (e.g., Bassett and Koenker, 1978), semiparametric analysis of binary response (e.g., Manski, 1985), semiparametric analysis of censored data (e.g., Powell, 1984) and semiparametric efficiency analysis (e.g., Chamberlain, 1986, 1987). See, for example, Kim and Pollard (1990) and Pollard (1990).

WHAT MAKES US DISTINCT?

Clogg stresses the commonalities between sociological methodology and statistics. It is also important to understand what makes these fields distinct. I do not know what Clogg sees as the factors that distinguish sociological methodology and statistics. I can, however, offer some thoughts on the distinction between econometrics and statistics. In particular, I shall contrast the way these disciplines have approached a subject of widespread interest, the analysis of binary response. The present discussion elaborates on remarks made in Manski (1988).

Suppose that one observes a random sample of observations on (z,x), where z is a binary outcome and x is a vector of regressors. Statistical analyses of binary response have aimed to *describe* the conditional probabilities $P(z = 1|x)$ as a function of x. The most general approach, and in some respects the simplest, is to regress z on x nonparametrically, using a kernel, nearest-neighbor, smoothing-spline or any other sensible method. Or one may fit a convenient parametric model, as is done in logistic regression.

Econometric analyses of binary response have aimed to *interpret* the conditional probabilities $P(z = 1|x)$ in terms of threshold-crossing models of the form

$$\begin{aligned} P(z = 1|x) &= P[f(x) + u > 0|x] \\ &= \int 1[f(x) + u > 0]dP(u|x) \end{aligned} \tag{1}$$

where u is an unobservable random variable and $P(u|x)$ is the distribution of u conditional on x. In economic applications, z usually indicates a decision maker's choice between two alternatives and $f(x) + u$ gives the difference between the utilities of these alternatives. The threshold-crossing model is also applied routinely in medical research. There z denotes an observable binary indicator of health status and $f(x) + u$ is a latent continuous variable determining health status.

The model (1) is vacuous in the absence of prior information on $f(\cdot)$ and $P(u|x)$. The econometric literature has determined the implications of a range of alternative assumptions. Semiparametric and parametric specifications are surveyed in Manski (1988); more recent work on nonparametric threshold-crossing models appears in Matzkin (1992).

I have observed that statisticians are often uncomfortable in the presence of these econometric models. They ask why it is not enough to describe $P(z = 1|x)$. Why should one want to interpret binary response as the outcome of a hypothetical threshold-crossing process involving a latent variable $f(x) + u$? I can offer three replies, of which the third is by far the most compelling to me.

1. Inference on economic concepts: Econometric models of binary response are formulated in terms of basic concepts of economic theory – preferences, expectations and opportunities. Empirical inference on these concepts is important to economic science. I don't expect statisticians to share this rationale for econometric modelling.

2. Precision of estimation of $P(z = 1|x)$: Suppose one is interested only in description, not in economics. If the prior information about the binary response process expressed in econometric models is correct, then invoking this information will, in general enable more precise estimation of $P(z = 1|x)$. The benefit may be an improvement in the rate of convergence relative to that obtainable nonparametrically or simply a reduction in asymptotic variance. Everyone – statistician or econometrician – likes more precise estimates. Of course, obtaining a gain in precision presumes that the prior information is correct.

3. Extrapolation: Nonparametric regression, invoking only local smoothness conditions, is capable of determining $P(z = 1|x)$ only on the support of the regressor x. But a great deal of econometric research is driven by the practical need to forecast the response z in a new economic environment, say one following a tax change or a technological change. Often, such a change can be represented as a change in the regressor values. Very often, the new economic environment carries x off its historical support. To extrapolate off the support of x, one has no alternative but to combine the available data with prior information. It may be that econometricians are foolhardy to attempt extrapolation, but, as economists, we are committed to doing so.

SOCIOLOGICAL METHODOLOGY AND ECONOMETRICS

Early in his article, Clogg suggests that analogous articles could be written from the vantage points of psychometrics and econometrics, but that he does not know enough about those areas to include them very much in his own article. A statistician with no background in the social sciences might be surprised to read these statements. Can it be that methodology varies so much within the social sciences that even someone as knowledgable as Clogg only feels comfortable discussing sociological methodology?

I understand Clogg's demurral fully. I would find it as difficult to write on the statistical contributions of psychometrics and sociological methodology. Clogg's article helped me understand not only the relationship between sociological methodology and statistics but also the relationship between sociological methodology and econometrics.

Econometricians and sociological methodologists have worked closely in some areas. In the early 1970s, they collaborated in the development of the latent variable models discussed by Clogg. See Goldberger and Duncan (1973). During the 1980s, both groups contributed to the development of rich models for the description of event-history data. Lancaster (1990) is a comprehensive and readable econometric treatment of the subject.

In other respects, econometricians and sociological methodologists have gone their separate ways. I was struck by Clogg's close association of categorical data analysis with the log-linear model, because the approach to discrete response analysis that took hold in econometrics during the 1970s was at most marginally influenced by the contemporaneous work on log-linear models. See the discussion in Manski and McFadden (1981).

I was also struck by Clogg's discussion of the survey-sampling literature on complex sampling, because this work has had essentially no impact on econometrics. Instead, we have developed the literature on estimation under "choice-based" sampling. The article by Hsieh, Manski and McFadden (1985), a survey written explicitly for a statistical audience, synthesizes this work and explains its relation to the biometric literature on "case-control" sampling. It is gratifying to be able to report that this effort at communication across disciplines has had some success. See, for example, Breslow and Cain (1988), who summarize and extend aspects of the econometric literature.

The point is that the various methodological disciplines form a complex social network, with strong relationships in some dimensions and weak ones in others. I find that econometricians and sociological methodologists speak much the same language on some subjects but can barely converse on others. I observe different mixes of the familiar and the strange when I read journals in psychometrics, biometrics and statistics. The various methodological disciplines have important shared foundations and objectives. But each one also has distinctive concerns which will, I suspect, keep them from coalescing any time soon.

ACKNOWLEDGMENT

I am grateful to Arthur Goldberger for his comments.

Comment: The Fence Between Statistics and Social Research

Ivo W. Molenaar

CLOGG IS RIGHT

Clogg (1992) defends the thesis that developments in sociological methodology and in quantitative sociology have always been closely related to development in statistical theory, methodology and computation. His impressive list of examples, from Quetelet's "average man" to event history analysis and finite mixtures, shows that he is right. It also shows that there was not just a relation between sociology and statistics: influences from quantitative researchers in psychology, education, economics, biology, demography, political science and management science can be found in many of Clogg's examples, and are indeed recognized by him. The development of covariance structure models, listed in his section "Models for Continuous Latent Variables," is an excellent example of how the concepts and skills of psychometricians, sociometricians, econometricians, statisticians and computer scientists were successfully brought together.

To the many examples cited by Clogg, I should like to add two: the adquate handling of missing observations, and the development of statistical computer packages.

Ivo W. Molenaar is Professor of Statistics and Measurement on the Faculty of Psychology, Sociology and Education at the University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.