

methodology but that it would also stimulate statisticians to make improvements so that it can bear comparison with the best statistical practice.

These, and many examples like them, might well be held to support the proposition that although social

science has produced a considerable amount of statistical methodology it has not sufficiently influenced the thought or direction of mainstream statistics. If this discussion helps to bring that about it will have been well worthwhile.

## Comment: It's the Interplay That's Important

Paul W. Holland

While he hides it well behind the mask of scholarly indifference, Clogg is hopping mad. He has had it up to here with effete mathematical snobs (couldn't data analyze their way out of a paper bag) telling him that, at bottom, all the really good ideas, even in statistics, come from mathematics and the physical sciences. The last straw was reading the same bilge in, of all places, R. A. Fisher! An understandable fury, but two things that it is good for applied statisticians to develop are a tolerance of foolishness and a very thick skin.

I have been "doing statistics" in the social sciences for most of my life, but on occasion the opportunity to examine raw data from the physical sciences arises, and it has always struck me on these occasions just how familiar they appear, even though they hail from an allegedly distant part of the scientific landscape. At the forefront of research, the high signal-to-noise ratio that some associate with physical science data simply isn't there. If it is, then we aren't seeing the new stuff, just the well-established old stuff. What does seem to distinguish the social from the physical sciences is that, in the latter, progress is always being made in instrumentation and the noise levels eventually get lower. Signals eventually stand out with great precision, but when they do, it is old news; it's just Kuhn's "normal science."

Most social scientists get used to the fact that, generally speaking, the noise level does not decrease (increasing  $n$  just introduces new ways to slice up the data, and we are then often back to our noisy little samples again). In this respect, progress in the social sciences is hard to make. Learning to live with this fact is part of the early socialization of social scientists and of the statisticians who work with them. Most

learn the hard way that a correlation over 0.8 is probably an error, and one exceeding 0.95 is an error for sure. (These limits must be modified somewhat for those who routinely correlate a variable with a slightly modified version of itself, but the same basic fact must be learned even there.)

The ubiquity of noise is why statisticians and statistics are useful to science—if there is no noise, no uncertainty, then there is no need for us or it.

*Oh Statist! seek Unruly's feast,  
and shun Perfection's meagre fare.*

Human variability is a root source of the wide application of statistics to social, behavioral and medical science, and the lack of such a reliable source of noise is why statistical science has a more limited role in routine, empirical, physical science.

Aside from these comments, I have little to say about the endless and, to me, the totally sterile "social versus physical science" debate. I do have opinions about some of the points and examples that Clogg mentions in his interesting paper, so I'll concentrate the rest of my comments on them.

There is no question in my mind that real problems, based on real data, influence the development of statistical procedures. How can it be otherwise? Example: structural zeros in multi-way contingency tables. Categorical variables can have impossible combinations—for example, male hysterectomies. Ignoring the fact that there will never be a non-zero entry in such a cell can lead to a false impression of the association in such tables, but I doubt if anyone ever thought much about the implications of structural zeros until they saw them in real data. However, structural zeros have long appeared in continuous distributions, for example, the uniform distribution on the unit disk, but not much is usually made of this, essentially mathematical, example except to counter the proposition that a zero correlation implies independence. This is a continuous example of the "quasi-independence" that Clogg discusses because the two variables are as independent as they

---

*Paul W. Holland is Distinguished Research Scholar at the Educational Testing Service and Fellow at the Center for Advanced Study in the Behavioral Sciences. Address correspondence to ETS, 21-T, Rosedale Road, Princeton, New Jersey 08541.*

can be, given that there are structural zeros. Structural zeros are an example of mathematics following data, but partly anticipating it as well. It's the interplay that's important.

Now a little history. An early major application of log-linear models for multi-way contingency tables was directed by Fred Mosteller in his contribution to the National Halothane Study which took place from 1963 to 1969 (Bunker et al., 1969). Yvonne Bishop developed her thesis out of that work which involved the full panoply of log-linear model fitting (estimation by maximum likelihood, structural zeros, likelihood ratio goodness-of-fit tests and the use of Freeman-Tukey residuals to assess model fit). This was a complex study in which notable statisticians (including J. Gilbert, L. Moses, F. Mosteller and J. Tukey) either directed or performed data analyses in which death rates associated with various anesthetics were adjusted in several ways for a variety of patient characteristics. Mosteller and Bishop were certainly aware of Goodman's work on log-linear models at that time (he is listed in the Foreword of the final report as one of *several* "statistical colleagues" who had been "generous with their advice") so this *might* be construed as one of Clogg's examples of the influence of sociological research on statistical and biometric research. Of course, at the time it was probably viewed, by the parties involved, as various statisticians working on related problems in different fields. It should also be mentioned that Birch (1963) was very influential on this work. The Halothane study had a strong influence on statistical research at Harvard, a direct result of which was the book by Bishop, Fienberg and Holland (1975).

For several years after the study, Bishop's computer program for fitting log-linear models by iterative scaling was the main computational tool available to interested researchers. Furthermore, I would not be the first in line to congratulate the developers of statistical packages for their help in disseminating the software necessary to fit these models. For years after the Halothane study the packages were great at noniterative linear methods involving nothing more complicated than a matrix of sums and cross-products, but they could not even *manipulate* a three-way array of counts let alone *fit a model* to one. To the statistical packages, categorical data just meant two-way tables and an ever expanding list of measures of association. The notable exception was the BMD package that always seemed to lead the others, but even BMD took quite a while before log-linear models were standard fare. Because BMD was heavily influenced by the work of biostatisticians, while the other packages were more influenced by the work of survey takers and psychologists, I wonder just how accurate Clogg is in his view of the influence of all the survey work on the further development of log-linear models. Of course, all this methodology is

now a routine part of any good, general statistical package.

It is interesting that Clogg chose to emphasize the concept of a *latent variable* as one without which "it is impossible to appreciate modern quantitative sociology." I think that his definition of a latent variable is much too loose. It is not simply that a latent variable is one that can't be measured directly. If you think about it, most variables are hardly ever observed. Rather, a latent variable is *unobservable in principle*, but it influences the values of the variables that can be observed, the *manifest variables*, in important and powerful ways. The whole purpose of a latent variable is to "explain" certain aspects of manifest variables. This notion of explanation is worth mentioning. Early on it was observed that if one obtains responses to a series of questions all bearing on a single theme (e.g., a series of vocabulary questions or a series of attitude questions about race relations) the responses to these questions all tend to be positively correlated. A latent variable named "vocabulary skill" or "degree of prejudice" is then hypothesized to explain these correlations. Such explanations can often be quite parsimonious; that is, using just a few parameters a whole matrix of correlations can be accurately reproduced. Formally, latent variable models always assume two things. First, given the values of all the latent variables, all of the manifest variables are statistically independent; that is, all their (conditional) correlations vanish. Second, the conditional distribution of any manifest variable given the latent variables depends strongly, but in simple ways (often linearly or at least monotonically), on the latent variables. The form of this dependence on the latent variables varies with the type of application. It is important to remember that without the second part of a latent variable model the first part, though apparently a strong assumption, is, in fact, completely vacuous in the sense that it places no testable restrictions on the data (Suppes and Zanotti, 1981; Holland and Rosenbaum, 1986).

Latent variable models were originally designed to "explain" the *associations* among a set of manifest variables rather than the *values* of these variables themselves. Some of the models that Clogg describes, and seems to extoll, go beyond these modest attempts to explain a set of associations and begin to relate various latent variables to each other. The error of reification is to believe that a name necessarily refers to something in the real world, even if the name is simply the product of an active imagination. For some, latent variables are even "more real" than the observed data that they are created to explain. For some, latent variables can be connected to each other by "causal" relations. They cast some of the shadows that dance on the wall of Plato's cave. In the theory of mental tests, a "true score," the name of the latent variable, is

sometimes regarded as a better measure than mere observed test performance, because it has been “purged” of measurement error. I have found some of the enthusiasm for latent variables difficult to understand because, from a technical point of view, all that can ever be known about the value of a latent variable are some aspects of its posterior distribution in some population given the manifest variables. In principle, the values of latent variables can’t ever be known. In the case of mental testing, it often happens (for long tests) that the posterior distribution of the latent variable is very nearly concentrated on a single point. In this case we don’t go far wrong by pretending that the distribution is a point mass. However, in most sociological examples, there are only a few indicators of the latent variable, and consequently the posterior distributions exhibit substantial variation. There are often barely enough relevant manifest variables to identify the model, that is, to be nonvacuous. I have very mixed feelings about the real value of latent variable models. On the one hand, I’m all for models that capture intuitions about data and that parsimoniously fit the details of real data sets, and it would be misleading if I gave the impression that these models never do either. On the other hand, so little separates these models from simply being vacuous that I think it is easy to overlook this fundamental point and to make more out of them than they warrant. The latent class models that Clogg mentions illustrate two related points. The first is what I call the discrete–continuous duality of latent variables. It is probably impossible in principle to decide whether a latent variable is continuous or discrete. It is not unusual for identical fits to data to be obtained by a latent class model and by another model that has a continuous latent variable. This is not all that surprising because finite mixtures often span or, at least, are dense in a space of infinite mixtures. Second, in the original use of latent class models, nothing was assumed about the relationship between the latent variable and the manifest variables except what I’ve called the first part of a latent variable model—the conditional independence part. The only reason that these models were not vacuous, that is, could not represent any set of data exactly and in more than one way, was the fact that the latent variable was assumed to be discrete, taking on only two or three values. It is almost an accident that such models are identifiable, and if the number of values of the latent variable is allowed to be a free parameter, then they cease to be. I think that good advice to a user of latent models is to “make latent variables powerful,” in the sense that they should have as many empirical consequences as possible, even to the point of trading off model fit for model simplicity. This is contrary to much of the sociological use of these models, in which a better fitting model might be achieved by weakening

the assumptions about the correlations between the unobservable latent variables.

I want to make a final point about latent variables, but before I do I need to make some preparatory comments about Clogg’s discussion of causal inference. It is, to me, a mistake to view the approach to causal inference in nonexperimental research advocated, say, in Rubin (1974) as *essentially* different from that which lies behind the causal model literature, for example, Blalock (1962). Clogg seems to say that they are essentially different. The data are the same, and the goal in both is to estimate causal effects. Elsewhere (Holland, 1988), I attempt to show the connection between these two approaches in a problem that involves both “direct” and “indirect” causation. In my opinion, Rubin’s approach is more detailed and principled than the causal model approach that begins directly with “error term” models. Sometimes, error term models can be deduced from Rubin’s approach, and then causal effect parameters can be identified with certain coefficients in the equations of these models. In other cases, it may be shown that the coefficients of a linear model do not have a causal interpretation. An important guiding principle of Rubin’s approach is to use comparative experiments as a model for thinking about problems of causal inference in nonexperimental studies. This leads us naturally to ask: What can be a cause in such a study? More precisely, what can have a causal effect? The answer is that if it can’t even be thought of as a treatment in a hypothetical comparative experiment then it is probably not a good idea to talk about its causal effects. In sociological studies, gender and ethnicity are usually non-causes in this sense, while education and occupation probably do have causal effects. One must always be careful in nonexperimental research to distinguish our ability to *define* causal effects from our ability to *measure or estimate* them in a specific case.

This brings us back to latent variables. Can a latent variable ever be a cause? In the structural model literature, this is no problem. Just define it to be a cause by fiat and adjust the model accordingly. But could a latent variable ever be a treatment in a hypothetical experiment? Because they are unobservable in principle and only their posterior distributions given the manifest variables can be known, they don’t seem to me to be good candidates for causes. The key question is whether or not it makes sense to imagine what a subject’s outcome would be if the value of the latent variable were one thing and to imagine another, potentially different, value of the outcome (for the same subject) if the value of the latent variable were something different. I don’t believe that this is the same thing as an experiment in which there is measurement error in the value of the treatment to which each unit is exposed, but in the structural model literature a

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.*