THE PECKING ORDER

Clogg is not very explicit about his reasons for writing his paper. One could wonder whether it is fruitful and appealing to stage a contest between disciplines about who contributed most to the development of statistics. Throughout the paper I perceive a slight irritation that the contribution of the social sciences is denied or belittled by statisticians. This is surely incorrect for those who play the favorable roles in Clogg's examples. It is probably correct for others, but I do not know for how many.

Clogg's paper can certainly help to enlighten those who were simply unaware of the developments cited by him. It remains to be seen, however, whether those newly enlightened statisticians, as well as those who already knew about the developments, will change their views.

This brings me to what I consider to be the latent structure, or hidden agenda, of the paper. This is the academic pecking order of disciplines. Even within mathematics, abstract topology and functional analysis are generally perceived to have a higher status than statistics. Within the domain of mathematical statistics, *The Annals of Statistics* enjoys a higher reputation than, say, *Biometrics* or *Psychometrika*. Among disciplines, mathematics, physics and biochemics have a better image among the outsiders than sociology or psychology.

The soft sciences cannot boast of spectacular achievements like sending astronauts to the moon or giving a patient another heart. In their relatively short history as academic disciplines, they have made less visible progress in areas like personnel selection, teaching methods, structure of organizations, ethnic tension and deviant behavior. Perhaps the nonbelievers in the usefulness of systematic empirical research in the social and behavioral sciences form a majority. They are found both within these disciplines, and in the ranks of the "harder" sciences. It may well be that in both camps one shares a feeling that a positivistic research style is not suitable in the study of human beings.

Formal models, of course, never catch the full richness and variation of human behavior and human feelings. It is equally true, however, that formal models of mechanics never fully catch the movements of real objects. In both cases, the abstract model can only be an approximation. Everybody knows and accepts that engineers can work with such approximations. In the study of human behavior it is rather more common to detect feelings of "it cannot be done, and, even if it could, for ethical reasons it should not be done." The book by Bartholomew (1973, section 1.3) contains some interesting thoughts on this issue in the context of applying stochastic process theory to social phenomena. Unlike the correctness of proofs or computer programs, this is an area where each individual has a personal value system that is seldom changed by discussion. Nevertheless, it may be useful to sometimes reflect on such matters, and the stimulating paper by Professor Clogg gave me an opportunity to do so.

Rejoinder

Clifford C. Clogg

I thank the discussants for their stimulating comments, many of which I judge to be consistent with the themes in my paper. The discussants cover several areas of statistical methodology that I either neglected or did not emphasize enough, provide more evidence for the claim that the context of social research has had a major effect on statistical methodology and give alternative points of view concerning how particular methodologies have developed. I agree with almost all of the points they make and so will confine myself to just a few remarks.

BARTHOLOMEW

I strongly agree that the social sciences place new demands on statistical methodology, particularly in areas such as measurement and measurement error, modeling correlated observations and latent variables. Bartholomew is right to refer to multiple correspondence analysis and recent advances in sampling theory as cases in point. The contrast between the natural or hard sciences and the social or behavioral sciences, insofar as statistical methodology is concerned, is very important to both his and my arguments. I tried to contrast the natural science setting with the social science setting a bit in my paper; also see Clogg and Dajani (1991).

HOLLAND

I was not hopping mad when I wrote the paper, but it is true that my tolerance for foolishness is so low that it could not be estimated with the fanciest of probit models.

Holland and I obviously have slightly different points of view on the development of the log-linear model. I tried to give a balanced view of this controversial subject and would be quite amenable to placing equal mass on each of the three main support points (Chicago, Harvard, Chapel Hill), with more than a little left over to spread around on others.

Holland's points about so-called causal models, the logical status of latent variables and whether variables can be causal need to be studied carefully. Holland is both an expert and an iconoclast on these subjects. Holland is right to ask serious questions about the principles of causation that may or may not apply to the models (and computer programs) that form the backbone of much contemporary methodology for social research. It seems to me that our best methodologists are taking Holland's arguments about these matters very seriously at the present time.

We do not know how to measure intelligence, but we all think that there is such a variable. Until the human genome project tells us otherwise, it is reasonable to think of intelligence as a latent variable and to act accordingly when we develop statistical models. Now surely intelligence is a causal variable for some outcomes; for example, intelligence is one of the many causes of social mobility. Social scientists are not so interested in how the measurement of intelligence with latent variable models relates to Plato's theory of forms. But they are very interested in designing better ability tests and checking results in "causal models" that incorporate notions of measurement error and summarize how either a social scientist or a policy maker tends to think.

MANSKI

I very much appreciated Manski's discussion of econometrics and statistics. That is an area that I should have said much more about, in part because econometric thinking is so commonplace in sociology and demography as well as in economics. His point that "the various methodological disciplines form a complex social network, with strong relationships in some dimensions and weak ones in others" is undoubtedly correct.

Many quantitative sociologists lean more on econometrics than on statistics, if I may be permitted to separate the two. Why is this the case? Manski hints at the answer when he explains why economists prefer econometric models over all-purpose tools for data analysis taught so much in statistics. In short, econometric models (or sociological models borrowed from econometrics) are designed to answer economic questions (or sociological questions), and they do so by bringing prior knowledge or existing theory to bear on the problem of fitting models to data. (This prior knowledge need not be of the Bayesian variety, but of course much of our best social research uses prior knowledge and is therefore Bayesian in some sense, sans the messy integrals.) Many of the standard methods taught in statistics courses on data analysis are not used so much in social research because they answer questions that are only moderately correlated with the questions that researchers ask.

MOLENAAR

Molenaar adds the handling of missing observations and the development of statistical software as key examples where the social science effect (in terms of individuals involved, scientific questions asked or data-analysis needs) has been "remarkable." I agree wholeheartedly.

I appreciate also Molenaar's comments on the philosophical underpinnings of statistical modeling in social science. Readers of this journal might not realize that statistical thinking is still not universally endorsed in sociology or in social research. Statistical modeling is often associated with logical positivism, and the latter is not a politically correct term in some circles. The association is so strong at the present time that those who criticize positivism are usually criticizing the idea that statistical models can be applied to human or social behavior. On this matter I would like to say for the record that the complement of positivism is negativism. A rigorous proof of this assertion can be done without measure theory. Readers not interested in the proof would do well to evaluate the output of social research done without statistical foundations. That stuff has high variance – perhaps mean-squared error is the better term – and progress is hard to spot.

Molenaar asks a very hard question: Why did I write the paper? I spent 7 years editing something or other, and the same question has popped into my mind more than once, but always about someone else's paper! I did not have a bad day in Sunday School. My main motive was not so much a concern with the pecking order, nor did I have anything special to contribute to the endless debate about natural or hard sciences versus social sciences. Rather, I think statisticians should pay more attention to the interplay among statistics, mathematics *and* the sciences, including the social sciences. My paper was about the importance of the interplay, to borrow terminology from Holland. I think most statistics graduate students go through their programs having studied about 10 books and perhaps a score of articles in statistics journals. They simply do not have much of a chance to see the interplay, except that with mathematics, much less appreciate it. This is a sad state of affairs, and it ought to be corrected.

FINAL NOTE

I sit on the fence between the social science areas with which I am familiar (sociology and demography) and statistics. The discussants are in similar positions; they just sit on the fence on different sides of the statistics field. They are on higher posts than I: they sit on the corner posts. It is gratifying to learn that other fence sitters have viewpoints that are generally consistent with my own. The main conclusion to draw from my article and their comments is that statisticians in the middle of the field would do well to come to the fence for greener grass. And I wish that more social scientists would at least look through the fence. Their failure to do more of that deserves at least as much commentary. The interplay among fields is very important for all concerned. Finally, I think we would all agree that it is fun to sit on the fence.

ADDITIONAL REFERENCES

ANDERSON, T. (1991). Trygve Haavelmo and simultaneous equation models. *Scand. J. Statist.* **18** 1–19.

ANDERSON, T. E. and RUBIN, H. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **5** 111–150. Univ. California Press, Berkeley.

BARNETT, W., POWELL, J. and TAUCHEN, G., eds. (1991). *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge Univ. Press.

BARTHOLOMEW, D. J. (1973). *Stochastic Models for Social Processes.* Wiley, London.

BARTHOLOMEW, D. J. (1987). *Latent Variable Models and Factor Analysis.* Oxford Univ. Press.

BASSETT, G. and KOENKER, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73** 618–622.

BEGUN, J., HALL, W., HUANG, W. and WELLNER, J. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11** 432–452.

BICKEL, P. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BRESLOW, N. and CAIN, K. (1988). Logistic regression for two-stage case-control data. *Biometrika* **75** 11–20.

BUNKER, J. P., FORREST JR., W. H., MOSTELLER, F. and VANDAM, L. D., eds. (1969). *The National Halothane Study.* National Institutes of Health, Bethesda, Md.

CHAMBERLAIN, G. (1986). Asymptotic efficiency in semiparametric models with censoring. *J. Econometrics* **32** 189–218.

CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34** 305–334.

CLOGG, C. C. (1992). The impact of sociological methodology on statistical methodology (with discussion). *Statist. Sci.* **7** 183–207.

FELLER, W. (1957). *An Introduction to Probability Theory and Its Applications.* Wiley, New York.

FIENBERG, S. and ZELLNER, A., eds. (1975). *Studies in Bayesian Econometrics and Statistics.* North-Holland, Amsterdam.

GADOUREK, I. (1982). *Social Change as Redefinition of Roles.* Van Gorcum, Assen.

GIFI, A. (1990) *Non-Linear Multivariate Analysis.* Wiley, New York.

GOLDBERGER, A. and DUNCAN, O. (1973). *Structural Equation Models in the Social Sciences.* Seminar Press, New York.

HAGENAARS, J. (1990). *Categorical Longitudinal Data.* Sage, Newbury Park, Calif.

HECKMAN, J. and SINGER, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52** 271–320.

HOLLAND, P. W. (1988). Causal inference, path analysis and recursive structural equations models. In *Sociological Methodology 1988* (C. C. Clogg, ed.) 449–484. American Sociological Assocation, Washington D.C.

HOLLAND, P. W. and ROSENBAUM, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14** 1523–1543.

HOOD, W. and KOOPMANS, T., eds. (1953). *Studies in Econometric Method.* Wiley, New York.

HSIEH, D., MANSKI, C. and MCFADDEN, D. (1985). Estimation of response probabilities from augmented retrospective observations. *J. Amer. Statist. Assoc.* **80** 651–662.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

LANCASTER, T. (1990). *The Analysis of Transition Data.* Cambridge Univ. Press.

MANSKI, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *J. Econometrics* **27** 303–333.

MANSKI, C. (1988). Identification of binary response models. *J. Amer. Statist. Assoc.* **83** 729–738.

MANSKI, C. and MCFADDEN, D. (1981). Alternative estimators and sample designs for discrete choice analysis. In *Structural Analysis of Discrete Data with Econometric Applications* (C. Manski and D. McFadden, eds.) 2–50. MIT Press.

MATZKIN, R. (1992). Nonparametric and distribution-free estimation of the threshold crossing and binary choice models. *Econometrica.* To appear.

MOLENAAR, I. W. (1988). Formal statistics and informal data analysis, or why laziness should be discouraged. *Statist. Neerlandica* **42** 83–90.

PAKES, A. and POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57** 1027–1057.

POLLARD, D. (1984). *Convergence and Stochastic Processes.* Springer, New York.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* IMS, Hayward, Calif.

POWELL, J. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics* **25** 303–325.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.

SKINNER, C. J., HOLT, D. and SMITH, T. M. F., eds. (1989). *Analysis of Complex Surveys.* Wiley, New York.

STONE, C. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

SUPPES, P. and ZANOTTI, M. (1981). When are probabilistic explanations possible? *Synthese* **48** 191–199.