Annals of Economic and Social Measurement 2/4, 1973

NBER COMPUTER RESEARCH CENTER NOTES

The NBER Computer Research Center for Economics and Management Science has been engaged, since its formation in 1971, in developing new software systems for quantitative social science research. Prototype systems for exploratory data analysis, mathematical programming, and econometrics are now in various stages of design and implementation. Notes on research in progress, as well as abstracts of working papers, are a regular feature in the Annals. *Following are a note on the robust regression study of the data-analysis project, and abstracts of three recent working papers (two of which present preliminary findings of the regression study).*

The Data-Analysis Project: Robust Regression Study (September 1973)

In our last note on the Center's data-analysis project (Holland, 1973), we defined our interest in resistant techniques for fitting linear models—with particular emphasis on robust regression and multiparameter Bayesian methods. Late last spring, we decided that the time was ripe to begin a comparative Monte Carlo study of robust regression methods. The purpose of the study is to bring together a variety of methods for fitting linear models, and to provide a broad base for evaluating not only the methods themselves but also different algorithms for the same methods.

While a Monte Carlo study is never a complete replacement for analytic results, it can be useful for weeding out poor methods and detecting promising ones. It will reduce the need for exact theory, but at the same time it will suggest interesting theoretical questions—in essence to "explain" the findings of the Monte Carlo study. One area of research where we expect to make contributions is the application of modern data-analytic techniques to interpret and clarify the results of Monte Carlo simulations of complex phenomena.

The feasibility and usefulness of this approach were established by the Princeton Robustness Study (Andrews *et al.*, 1972). Not only did they succeed in providing a useful comparative study, but the work led to new methods and devices for combining methods to produce useful hybrids.

In order to gain feedback about the design of the study, and to bring together researchers interested in these problems for a concerted assault on them, the Center held a working conference on robust regression on June 21–22, 1973. The following scholars from outside institutions participated:

David F. Andrews
(University of Toronto)
Michael L. Brown
(Harvard University)
Gregory C. Chow
(Princeton University)
Ray C. Fair
(Princeton University)
George S. Fishman
(Yale University)
Joseph L. Gastwirth
(George Washington University)
Nan Hughes
(Harvard University)
Colin L. Mallows
(Bell Laboratories)
R. Douglas Martin
(University of Washington
and Bell Laboratories)
Frederick Mosteller
(Harvard University)

William H. Rogers
(Stanford University and
Bell Laboratories)
Barr Rosenberg
(University of California, Berkeley)
Stephen M. Stigler
(University of Wisconsin)
John W. Tukey
(Princeton University and
Bell Laboratories)

The following members of the Center staff participated:

Swarnjit S. Arora
David A. Belsley
Mark Eisner
Donald E. Farrar
Richard Hill
David C. Hoaglin
Paul W. Holland
Edwin Kuh
John R. Meyer
Roy E. Welsch

The study began in the summer of 1973 with the programming of a series of robust estimators, including:

(1) A few of the best location estimators from the Princeton study.
(2) Two robust methods, median polish and α-trimmed mean polish, for fitting an additive model to a two-way table. Both methods are alternatives to the usual analysis-of-variance techniques.
(3) Linear regression by minimizing the sum of absolute residuals. This is a beginning for many robust regression methods.
(4) Iteratively reweighted least squares in which the weights at each iteration are computed from the residuals of the previous iteration. This serves as a basis for computations in a variety of robust regression methods.

In addition, various random number generating schemes were implemented in anticipation of the needs of the Monte Carlo study.

The routines described above were programmed for interactive use within the Center's TROLL system and are available to users of the telecommunications network in which the Center participates through the NBER's Computer Operations Activity.

David C. Hoaglin
Harvard University
NBER Computer Research Center
Paul W. Holland
NBER Computer Research Center
Roy E. Welsch
Massachusetts Institute of Technology
NBER Computer Research Center

References

Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press.
Holland, Paul W. (1973). "The Data-Analysis Project (March 1973)". *Annals of Economic and Social Measurement*, Vol. 2, No. 3, pp. 221–224.

Abstracts of Working Papers

Working papers of researchers at the Computer Research Center are now published in the National Bureau's general Working Paper Series and are also abstracted here. The full text of the papers are

available in limited quantity, at $1.00 per copy, from the NBER Computer Research Center, 575 Technology Square, Cambridge, Massachusetts 02139 (Attention: Support Staff).

Arora, Swarnjit S. (University of Wisconsin, Milwaukee, and NBER Computer Research Center), "Error Components Regression Models and their Applications", NBER Working Paper No. 3 (June 1973), 24 pp.

In this paper, we have developed an operational method for estimating error components regression models when the variance-covariance matrix of the disturbance terms is unknown. Monte Carlo studies were conducted to compare the relative efficiency of the pooled estimator obtained by this procedure to (a) an ordinary least squares estimator based on data aggregated over time, (b) the covariance estimator, (c) the ordinary least squares estimator, and (d) a generalized least squares estimator based on a known variance covariance matrix. For T small and large ρ, this estimator definitely performs better than the other estimators which are also based on an estimated value of the variance-covariance matrix of the disturbances. For ρ small and large T it compares equally well with the other estimators.

Holland, Paul H., "Monte Carlo for Robust Regression: The Swindle Unmasked", NBER Working Paper No. 10 (September 1973), 15 pp.

Gives an alternative derivation of a Monte Carlo method that has been used to study robust estimators. Extensions of the technique to the regression case are also considered and some computational points are briefly mentioned.

——, "Weighted Ridge Regression: Combining Ridge and Robust Regression Methods", NBER Working Paper No. 11 (September 1973), 19 pp.

Gives the formulas for and derivation of ridge regression methods when there are weights associated with each observation. A Bayesian motivation is used and various choices of k are discussed. A suggestion is made as to how to combine ridge regression with robust regression methods.