Volume Title: Annals of Economic and Social Measurement, Volume 1, number 2

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm72-2

Publication Date: April 1972

Chapter Title: Some Notes on the Research Program at the NBER Computer Research Center for Economics and Management Science

Chapter Author: Edwin Kuh

Chapter URL: http://www.nber.org/chapters/c9196

Chapter pages in book: (p. 233 - 236)

Annals of Economic and Social Measurement, 1/2, 1972

PROGRAMMING SOFTWARE NOTES

SOME NOTES ON THE RESEARCH PROGRAM AT THE NBER COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE

BY EDWIN KUH

In broadest perspective, the objective of the Computer Research Center is to develop significant research tools using advanced system programming technology and to distribute these tools to researchers and the broader user community. There are several activities that might fall under this description which, however, do not. To clarify matters, I will mention three of the most important. First, the Center will only incidentally be concerned with the distribution of existing programs; rather, we shall be creating our own programs in a general systems framework particularly oriented to the needs of quantitative research in economics and management science. Second, the Center will not be concerned with highly speculative basic research. Most such work is, almost by definition, not well enough structured to have any reasonable certainty about its relevance to the user community. Third, insofar as our objective is systems development built around related algorithms or methods, the Center is not a grant agency designed to fund isolated projects.

Research objectives are established by the Center Policy Committee.[1] The research will be carried out in the following manner. The programming staff of ten full-time equivalents will be resident at the Center. Three or four academic researchers will be in residence to carry out their own activities directly through computer and programming facilities available at the Center, as well as to provide a communication link with seven university-based researchers. University-based researchers will be able to access the programs through a console using the new Center operating system, COS, now being designed for the IBMS/360–67. A number of university faculty will also be at the Center for a semester or two. A support staff of five persons will provide internal and external documentation as well as advice to external users about technical problems that arise in the applications of systems developed by the Center.

This summer we surveyed research and came up with recommendations for priorities. Two major topics were selected to be our research sphere for the next several years, mathematical programming and data analysis. Other topics include full-information maximum likelihood, non-linear estimation, Monte Carlo simulation, spectral analysis, and Box—Jenkins—Philips moving average error estimation.

[1] The members of the Policy and Operations Committee of the Center include W. J. Dixon, UCLA; Harvey Wagner, Yale; Robert Dorfman, Harvard; William Sharpe, Stanford; Thomas A. Wilson, Toronto; Gregory C. Chow, Princeton; Edwin Kuh; John R. Meyer, Yale and NBER; and T. C. Liu, Cornell.

Harvey Wagner at Yale University, assisted by Jeremy Shapiro at MIT, held discussions with twenty colleagues about prospects for significant achievements in mathematical programming within the research framework of the Center and what steps might be initiated to carry off an endeavor of this sort.[2]

Present knowledge of mathematical programming is so sophisticated that future progress requires a close and immediate working relationship between the development of mathematical techniques and their implementation in efficient computer systems. Today, the mathematical programming profession is wary of technical ideas that are advanced without any demonstration of computational effectiveness. The unique strength of the Center is that it can assemble an experienced staff of mathematical programming systems scientists, maintain this staff on a full-time basis for a sufficient period to produce worthwhile advanced systems, and provide the necessary research support (working environment, supporting personnel, and availability of large blocks of computing time on a large-scale machine).

Even in leading educational institutions where there are a substantial number of mathematical programming theoreticians and computer scientists, progress in the development of such systems software has been halting. Three prime reasons for this spotty record are the lack of full-time, experienced systems programmers available for a duration of 12–24 months, insufficient computer budgets to develop the systems, and a dearth of real problems to test the efficiency of the programs. Consequently, the Center has been established at an opportune time in the development of mathematical programming software systems.

Mathematical programming scientists for the most part agree as to which areas of systems research ought to receive highest priority. Differences concerning specific approaches, of course, do exist. Actually, the various research areas have system elements in common, so that regardless of the topics investigated initially, subsequent research will be well served by the early projects of the Center.

Specifically, the important areas of research include:

—Large-scale linear programming problems
—Large-scale non-linear programming problems
—Combinatorial and discrete programming problems
—Algorithmic languages for mathematical programming
—Innovative applications
—Methodology for comparing computational techniques.

The research activities of the mathematical programming staff can be expected to yield external economies and economies of scale both to the mathematical programming profession and to potential users of this technology. For example, although the Center itself will maintain only a limited number of advanced systems for solving various mathematical programming problems, the staff should have an up-to-date knowledge of systems available elsewhere. Hence, the Center should be able to advise researchers dealing with applications as to the best available systems for their problems.

[2] The following remarks draw on a September 20 memorandum prepared by Harvey Wagner for the Center.

Data Analysis

The problem of data analysis is as large as the problem of empirical research itself, so that one must consider those aspects which lend themselves to advances in the state of the art in an efficient way. Somewhat more narrowly conceived, data analysis is the process of integrating statistical theory with the realities of applied problems, which will often require intelligent heuristics. Since this research area is not tightly structured, its content will evolve in ways that cannot be clearly foreseen. Nevertheless, for our purposes, there are several related areas that will be considered initially.

In most current econometric research, a model is set up, data acquired, and then the coefficients estimated and interpreted. By way of contrast, data analysis places a substantial emphasis on visual and parametric exploration of the data series initially. Hence, although there normally is some loosely conceived model in mind from the outset, the data analysis approach de-emphasizes *a priori* restrictions.

Thus, preliminary emphasis is on visual presentation of data in ways that will clarify distribution characteristics, e.g., outliers, skewness and multi-modality. The Center will develop graphical capabilities to this end. This should not be difficult, although much care is essential in devising a natural language to suit this purpose. A draft text on *Exploratory Data Analysis* by John Tukey embodies a specific and rich empirical methodological approach to data analysis problems which statisticians at the Center could draw on for a variety of ideas useful to econometricians and management scientists.

The next step in the model building process from the viewpoint of data analysis is the systematic exploration of data to better understand its internal composition in a way that extends beyond the data descriptions described above. One promising approach is cluster analysis. This method, along with some others, should become part of Center research activity. Multivariate cluster analysis, where there is prior information about the variables influencing the entities of interest, can be approached in several ways. One of the most appealing is to have a cross-tabulation capability where the prior information is used in establishing characteristics for the cross-tabulation. Then cluster analysis is performed within cross-tabulated cells. Other aspects of multiple cluster analysis are still in their infancy, but it should be possible for the Center to extend this type of research in useful directions.

Since real world data often violates convenient assumptions about error processes assumed by mathematical statisticians, the development of robust estimators is an integral part of data analysis. Under Tukey's leadership, a group of statisticians at Princeton University has made an excellent analysis of robust estimators of the mean where the frequency distribution is non-normal. Econometricians could benefit greatly if these tools were readily available. There is other work now going on in the area of robust estimation of regression models. The Center is canvassing the possibilities of these alternative approaches to promote research and subsequent dissemination.

One problem that arises frequently in applied econometric work is stability of the underlying regression regime. Regression coefficients often change from one set of data to the next. When some discrete event like a war or natural disaster (or blessing) can be clearly identified, ordinary analysis of covariance is directly applicable. When, however, the point at which the break or breaks in the regime must be estimated along with the different sets of regression coefficients, the problem becomes both more realistic and difficult. Several approaches have been proposed including work by Richard Quandt and James Durbin. Since the problem is a pervasive one, the Center intends to promote research and implementation of algorithms on this topic.

National Bureau of Economic Research
and Massachusetts Institute of Technology