

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: <http://www.nber.org/books/aesm72-1>

Publication Date: 1972

Chapter Title: Computers, Information Retrieval and Research Methodology

Chapter Author: Sanford V. Berg

Chapter URL: <http://www.nber.org/chapters/c9179>

Chapter pages in book: (p. 1 - 5)

COMPUTERS, INFORMATION RETRIEVAL AND RESEARCH METHODOLOGY

BY SANFORD V. BERG

Computers have contributed to making economics a mathematical science, and are changing the very character of quantitative research. The old division of labor among researchers did not adequately promote the exchange of information about available data, software, and new research techniques. Because of this situation, the NBER has begun several projects, including the publication of a new journal, to reduce communications lags and extend recent advances.

The purpose of the *Annals of Economic and Social Measurement* is to serve as an interface between economics and statistics, management science, and information science. The modern computer has created a large constituency with common problems, requiring a new division of labor which permits more efficient specialization. Such specialization cuts across fields and reflects an increasing awareness of the problems and potential associated with large-scale research efforts—exchange of data, distribution of software, and dissemination of new methodologies for quantitative research. Although initial developments in research methodology need not necessarily involve the computer in a vital way, further improvements in measurement and estimation will probably have roots in computer technology.

THE INVISIBLE SCIENTIFIC REVOLUTION

During the last decade, the computer has contributed to the emergence of economics as a mathematical science.¹ Although the way research is conducted has been altered, models have been refined rather than abandoned.² This conceptual continuity has been attributed to imprecise theories and unmet data needs; the models are less susceptible to empirical testing.³ However, the storage and manipulative capacity of computers is being combined more effectively with advanced techniques, so mathematical models are becoming more powerful tools.

The computerized information retrieval systems appearing in universities and government statistical offices promise to further alter the research patterns of economists, just as the telescope and microscope changed the research methodology of astronomers and biologists. The change cannot help but have an impact on our socio-economic models, given the possibilities inherent in the new technology. The Behavioral and Social Sciences Survey for Economics,⁴ assembled by a distinguished panel of economists, placed great emphasis on changes being

¹ For a description of the non-computer aspects of the process, see Karl Fox, "The Invisible Revolution in Economics: Emergence of a Mathematical Science," in *Economic Models, Estimation and Risk Programming*, Springer-Verlag, 1969.

² George Stigler, "Does Economics Have a Useful Past," *History of Political Economy*, Vol. 1, No. 2, 1969.

³ A. W. Coats, "Is There a 'Structure of Scientific Revolutions' in Economics," *Kyklos*, Vol. XXII, 1969, Fasc. 2.

⁴ Nancy Ruggles, ed., *Economics* (Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1970).

wrought by the computer. The panel made six recommendations involving computers, information retrieval, and changes in research methodology:

- (1) the development of a national data system for economic and social research;
- (2) the establishment of large scale social science research centers to function as laboratories for the social scientist;
- (3) the provision of computer facilities for research, and the design of computer networks which can provide linkages between research centers and data bases;
- (4) the organization of long-term economic research efforts to reduce start-up costs, build trained staffs, and choose more efficient research methods;
- (5) the reorientation of graduate training toward providing more competence in research methods and more research experience, with particular emphasis placed on the acquisition of mathematical, statistical, and programming skills; and
- (6) the improvement of communication channels within the profession, and the sponsorship of workshops and conferences in specialized fields.

The entire Report, from which the above recommendations are distilled, stresses the changes in the character of research brought about by the availability of new technologies, like the computer and telecommunications, and new research methodologies, such as simulation, large-scale econometric modeling, and sophisticated estimation techniques. The potential for even more dramatic change is at hand.

COMPUTERS AND RESEARCHERS

After a period of intensive experimentation with computer technology and its application to a wide range of research activities, scientists are in a position to consolidate and extend recent advances. Consolidation is necessary because the fast pace of technological change has resulted in the rapid turnover of computer hardware and the duplication of programming efforts. The National Bureau has responded to the current situation by creating a *Conference on the Computer in Economic and Social Research*. The series of workshops held under the auspices of the Conference promotes the exchange of information and experience among scholars from universities, research institutes, government agencies and other groups concerned with large-scale computer research on economic and social questions. The *Annals* will publish papers on databanks, microdata sets, research methodology, and information retrieval which are presented at workshops.

Further extension of advances in computer technology is made possible through a second NBER project, a *Computer Research Center for Economics and Management Science*. Through the Center, researchers will interact with a skilled staff of programmers and systems analysts; the operating system will facilitate the creation, testing, and modification of new computational techniques. These endeavors should result in new software for a broad class of users. Programs will be disseminated through educational activities (such as workshops) and maintenance procedures, with the *Annals* serving as a clearinghouse for the Center.⁵

⁵ One NBER software package developed prior to the establishment of the Center is described in Y. Haitovsky and S. Jacobs. "REGEN-Computer Program to Generate Multivariate Observations for Linear Regression Equations," this issue.

Computer technology can be expected to further change our concepts of economic and social measurement. Within a decade, many economists will spend as much time working at interactive terminals as their predecessors did at their desks, with pencil and paper. The shift to the use of new research techniques is likely to be quite discontinuous and involve high adjustment costs. During the transition, wide dissemination of information about available programs or new data manipulation systems can reduce the communications gaps resulting from overspecialization and isolation.

The computer is already aiding scientists in both microeconomic and macroeconomic research; many instances will be reported in the *Annals*. The development of testable hypotheses has plagued researchers in microeconomics, but simulation models of consumers and producers offer promise for the future.⁶ The estimation and simulation of large-scale macroeconomic models is also made possible by the computer.⁷ In both cases, the econometrician is able to specify much more complex relationships than were previously feasible, with simulations and sensitivity tests becoming regular aspects of his research. Researchers are beginning to formulate models with sufficient complexity, that models more adequately reflect reality; software which handles complicated lag structures and nonlinearities has facilitated these efforts. However, the models often lack enough simplicity to be understood, or do not have unambiguous interpretations of casual relationships. Furthermore, the quality of basic data may not permit one to discriminate from among alternative hypotheses embedded within these sophisticated models.⁸ Problems of research methodology, estimation techniques, and data needs are not likely to diminish when the fourth generation of computers become widely available.

INFORMATION RETRIEVAL

Data accessibility must be insured in order for the computer to serve as an efficient tool. Problems of data documentation, storage, and retrieval will be a major focus of *Annals* articles. A forthcoming article on standards for data interchange is partially a result of a databank workshop sponsored by the inter-institutional Conference on the Computer. The standards developed by George Sadowsky and others provide for human readability; header systems that can be interrogated by standardized programs could provide labels, definitions, and source-notes for researchers. Hopefully the system will prevent much needless duplication of effort and ensure ease of access to data.

A second example of developments in information retrieval is the NBER databank of over 2,500 time series. This centralized repository of economic data

⁶ Martin Shubik believes that new developments in computation and simulation will result in the joining together of microeconomic institutional studies, advanced mathematical-economic theory and political economy. With computer technology, complex institutional elements can re-enter our market models: for example, game-theoretic considerations can be taken into account through the introduction of conjectural variations. Martin Shubik, "A Curmudgeon's Guide to Microeconomics," *Journal of Economic Literature*, June 1970.

⁷ See *Econometric Models of Cyclical Behavior*, Bert Hickman, ed., forthcoming, NBER.

⁸ One aspect of determining causation is the identification of endogenous variables. See Christopher Sims, "Are There Exogenous Variables in Short-Production Relations?" this issue.

was created and is being maintained for internal research purposes and for outside users who subscribe to the information service. The data base has become part of an on-going effort to trace the changing requirements for data processing in economics, to establish efficient methods for gathering and handling economic data, and to collect information about the technology of data retrieval. Readers of the *Annals* will be kept informed of developments in the NBER⁹ and other databanks.

As a supplement for book reviews, we hope to have reviews of major machine-readable data sets. For example, a forthcoming Conference on the Current Population Survey will generate papers by researchers within the Census Department and by outside users of CPS data. Papers on the objectives of CPS and conceptual problems with the data will be published in the *Annals*.

The computer can be expected to affect our capabilities to store and retrieve not only data, but also information about research. Information that has traditionally been disseminated through published media will begin to become available on-line. For example, the *Journal of Economic Literature* is experimenting with machine-readable input. Another NBER workshop has examined the role of the computer in the evolution of communication networks, with some papers to receive wider dissemination through the *Annals*.¹⁰ As more applications occur, we can expect to find an increasing dependence on the computer for information retrieval.

RESEARCH METHODOLOGY

The analytic techniques made possible by new technology alter the production possibilities for economic knowledge. Assuming that the past mix of activities adequately reflected relative prices and payoffs, a reallocation of resources devoted to research is necessary for efficiency. Tom Juster, in his paper in this issue, argues in favor of devoting more resources to the collection of data for the testing of hypotheses.¹¹ A shift of resources to research centers which are capable of recruiting, training, and supporting large-scale quantitative research is another change brought about by new technologies. There is also need for collaboration among a number of researchers trained in different fields, as was pointed out by the BASS Report discussions of survey research and simulation models.¹² The methodological advances resulting from such cross-fertilization will find the *Annals* a particularly appropriate communications channel.

In a way, the journal system operates as a kind of "invisible hand," aiding in the coordination of activities of diverse specialists. Journals alert scientists to new research possibilities, and reward (indirectly) those who publish research results. Such indirect coordination may break down in periods of rapid change. Methodological advances tend to be linked to substantive research, so disciplinary boundaries can impede widespread application of the advances. Future papers will illustrate how the solution to particular data problems can

⁹ Charlotte Boschan, "The National Bureau Time-Series Databank," *Annals of Economic and Social Measurement*, forthcoming.

¹⁰ See the note by Neville Beharie for an overview of the workshop.

¹¹ See F. Thomas Juster, "Microdata Requirements and Public Policy Decisions," this issue.

¹² Nancy Ruggles, *op. cit.*

help scientists in other disciplines.¹³ Researchers in health, education, and other fields are combining data sets from different sources with various degrees of success. Experimentation with alternative match/merge procedures can be expected in the coming years as more attempts are made to graft additional variables onto empirical models. Such efforts are in many ways more difficult than arm-chair theorizing, but are potentially more productive.

NEW DIRECTIONS IN RESEARCH AND THE NEW JOURNAL

Quantitative research has received a giant boost from the computer. This change is desirable, if we consider the possible development of economics without an infusion of new analytic techniques:

... there is grave danger that the subject will develop along the line of least resistance, that the stream so far from its source will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much abstract inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up ...¹⁴

Is the danger signal up in Economics? In his Presidential address to the American Economic Association, Wassily Leontief spoke disparagingly of the "state of splendid isolation" in which Economics finds itself and called for a shift to "systematic large-scale factual analysis."¹⁵ If Economics is slipping into a phase in which research is superficially impressive but lacking in substantive content, then developments in computers, data bases, and research methodology can serve as a powerful counterforce. We hope that the National Bureau, through the *Annals* and other projects, can focus attention on the role of the computer in large-scale factual analysis.

It is the hope of the board of editors that the articles published in the new journal will inform, instruct, and motivate researchers in economics, statistics, and management science. Thus, the *Annals* is intended to be more than a research archive for retrospective searching; the articles will also serve to increase awareness of current developments through the publication of timely articles and announcements. We will give visibility to researchers whose work has a high social marginal product, but who often do not receive the recognition warranted by their work. We feel that the articles carried in the journal will help express the sense that intellectual and technological breakthroughs in economic and social measurement offer exciting possibilities for future research.

*University of Florida, and
National Bureau of Economic Research*

¹³ Ben Okner, "Constructing a New Data Base From Existing Microdata Sets—1966 Merge File," forthcoming.

¹⁴ J. von Neumann, "The Mathematician," in R. B. Heywood (ed.), *The Works of the Mind* (Chicago: University of Chicago Press, 1947), p. 196.

¹⁵ Wassily Leontief, "Theoretical Assumptions and Nonobserved Facts," *American Economic Review*, March 1971, p. 5.