The International Journal of Organizational Analysis
1993, Vol. 1, No. 3 (July), pp. 255–272

MANAGEMENT RESEARCH METHODOLOGY: PROSPECTS AND LINKS TO PRACTICE

George T. Duncan
Carnegie Mellon University

Future developments in methodology have the potential to improve management research and better couple it to management practice. These developments are on six fronts: (1) computer technology, (2) data capture and experimentation, (3) privacy, confidentiality, and data access, (4) causation, (5) modeling and simulation, and (6) Bayesian statistics. The potential of each is explored, and problems, both technical and administrative, in fulfilling this potential are identified. On the computer and communications front, the key elements are the use of relational database management systems, increased computing power for analysis purposes, and computer networking. On the privacy, confidentiality, and data access front, the key elements are new capabilities for data capture through real-time surveillance, inferential disclosure threats in computer databases, the demand for more access to detailed data, and public concerns for privacy invasion. Management research is a search for causal mechanisms that can be investigated through empirical studies and that facilitate control of complex processes. In the modeling area, there will be (1) greater use of computing power, (2) less use of model-independent statistical hypothesis testing, and (3) easier to use computer software for modeling and simulation. The Bayesian perspective of consistently expressing uncertainty through probability distributions will become more widely used in management research.

"Progress . . . is the mother of problems." So said Gilbert Keith Chesterton, the versatile and iconoclastic English author. Progress on both substantive and methodological fronts deliver engaging problems for the management researcher. Changes in social values, organizational structures and management practices yield new substantive problems to address. These new problems are the stuff of management research, whether it be a search for an organization theory for a post-socialist post-industrial society, an examination of American joint ventures in Hungary, or a study of operational modes for multinational corporations in post-

---

Note: This work was partially supported by NSF Grant SES 9110512

apartheid South Africa. Methodological progress, especially that driven by increased computing power, also raises challenging new problems for the researcher.

Management research, like other areas of research, cultivates the ground of substance—the facts and theory of the problem at hand—with the spade of methodology—the techniques and perspectives of doing research. To avoid having efforts result in nothing but a weed patch, each productive management researcher must stake out some substantive ground and cultivate intensively with a methodological spade. This article focuses on how the spade of methodology must adapt to these progeny problems of progress. On six fronts there are developments that have genuine promise to both improve management research and better couple it to management practice. The following identifies these fronts and suggests the nature of the development. The rest of the article explores the implications of these developments.

First, computer technology: A May, 1993 *Byte* magazine displays a 486DX2 machine running at 66 megahertz with 8 megabytes of RAM, two floppy drives, a 340 megabyte hard drive, a 15" color monitor, 124–key keyboard, mouse, Microsoft Windows 3.1, a CD–ROM drive, and a choice of applications software like Microsoft Excel for Windows—all mail order for under $3,000.

Second, data capture and experimentation: How can we get some of the best features of experimentation while basically doing observational studies?

Third, privacy, confidentiality and data access: How do we better protect privacy and insure confidentiality of respondents, while at the same time assuring researchers better data access?

Fourth, causation: How do we infer causal relations?

Fifth, modeling methods and simulation procedures: Can new techniques advance our understanding?

Sixth, Bayesian statistics: Can Bayesian methods contribute to management research?

Impact of Computer Technology

The impact of computer technology is obvious to us all. It has profoundly affected the way we work. Yet the near future suggests further challenges and opportunities. Key elements of this are the use of database management systems, increased computing power for analysis purposes, and computer networking. An example of the kind of capability that will become more widespread is CDC WONDER/PC which is the Center for Disease Control's online public health information system providing remote access to large data bases, graphics and mapping, electronic mail, and surveillance and survey data uploads and feedback.

Relational database management systems

In relational database management systems, the user perceives the data as organized in tables with exactly one value in each row–column cell. Data manipulation occurs through operators on the tables. "Join" is one such operator that takes values from one table and matches them with corresponding values from another table. As an example, a government agency may have a database in which one table contains records for employees on participation in sponsored skill-enhancement programs and a second table contains employee salary and promotion records. "Join" could list all employees who had participated in a certain program along with, say, assessments of their performance matched with chosen aspects of their salary history.

What role will such relational database management systems play in management research? Although they were originally conceived as an operational tool in business and government (see Date, 1983), their potential for research and statistical use was recognized early (Chambers, 1977). Today, implementation difficulties have been overcome and relational database management systems such as Oracle, FoxPro and Paradox are beginning to play important roles in structuring research data files. Their value in areas of interest to management research is beginning to be recognized (David, 1989; Gates, 1989).

Using a relational database management system has several advantages. It substantially eases data editing. It allows interesting subgroups to be easily constructed. And it permits data from various sources to be combined through record matching. Database management systems have value for both primary analysis—where data collection is integral (see, e.g., Federer, 1973 for a conceptual and elementary treatment)—and secondary analysis.

With sophisticated database management systems it will be essential that their be adequate documentation of these dynamic archives. Some of the issues are addressed by Card and Peterson (1991), and David (1991).

Primary Analysis: Structure of the Computer Database

For primary analysis, researchers will spend more time on the content and design of the structure of the computer database. They will be more concerned with measurement issues (see, e.g., Krantz, Luce, Suppes, & Tversky, 1971; Roberts, 1979):

1. Is productivity a univariate concept?
2. What aspects of employee stress relate to organizational performance?

They will answer questions like:

1. How does a data file of employment records relate to a data file on electronic mail messages?

2. How should the database handle departments that are formed, are abolished, are merged, or renamed?

Researchers will increasingly want to collect complicated microdata with extensive context variables:

1. What is the educational and experience background of workers?
2. Who are their social and professional associates?

Secondary Analysis: Use of Database Query Languages

For secondary analysis, researchers will find extensive microdata files available both from private and public sources. They will have access to sophisticated and easy-to-use database languages (or query languages), which will easily allow the researcher to manipulate data according to the four basic operations of retrieval, modification, deletion, and insertion. They will probably be based on what is becoming the industry standard—SQL (for Structured Query Language). To some extent they will use natural language query, which will allow the English speaker to make queries like: "Which Regional Offices are Overbudget on Personnel Costs?" and the Japanese speaker to make the query: "Dono Ofice Ga Kajo No Jinken–Hi O Motte Imasu–Ka?" A start on natural language query in English has been made with the language INTELLECT. Further progress can be expected with (1) advances in linguistic analysis and (2) cheaper computer power to translate language into machine code equivalents.

Large Databases

Researchers will have substantially increased computing power to work with enormous databases. In working, for example, with administrative records, such as Federal Government Civil Service personnel records detailing promotions, job changes and job tenure, a researcher would have the computational power to work with the entire database, rather than a 1 or 10 percent sample of the database. A clear gain from this is more precise statistical information regarding aggregates, but another gain is the ability to do more refined subgroup analysis—say for women born in Mexico now employed by the U.S. Department of Housing and Urban Development.

Computationally-Intensive Statistical Methods

Researchers will also be able to employ sophisticated, computationally-intensive statistical methods. A revolution has occurred here. Consider the course of particular domains such as nonparametric statistics. Twenty years ago a useful topic was building "back-of-an-envelope" methods that might guide shop-floor quality control. Today, the shop floor is hardly neglected, because "industrial-strength" 486 PCs are right there to do computation for statistical quality management that is well beyond that possible on the back of an envelope. Yet, nonparametrics research has not died out. Instead, the emphasis now is on computationally-involved methods, taking advantage of the fact that "today's data analyst can afford to expend more computation on a single problem than the world's yearly total of statistical computation in the 1920's" (Efron & Tibshirani, 1991, p. 390). Illustrative of this are some topics in nonparametric statistics that are under continued methodological development:

1. kernel density estimators that can effectively construct smoothed estimates of probability density functions from large samples (see, e.g., Samiuddin & El-Sayyad, 1991),

2. bootstrap methods which seek to get confidence intervals using samples without replacement from the original sample (see, e.g., DiCiccio & Romano, 1988),
3. nonparametric regression methods which fit a sequence of local regression curves to moving windows of the data points (see, e.g., Cleveland, 1979).

Statistical techniques in each of these three topic areas require computing resources that would have been considered enormous a few years ago, but today most researchers have the necessary computer hardware, if not yet the software, on their desktops.

In a quite general sense, the researcher now has the computational power, using a personal desktop computer, to implement highly sophisticated tools and to develop new tools, with relatively few concerns about the cost of computation. This computing power will be most evident in permitting dynamic computer graphics—computer-generated pictures that evolve on the screen over time (Miller & Wegman, 1989; Hurley & Buja, 1990). For example, spins of three-dimensional data can give a remarkably more informative portrait of the joint distribution than can a variety of two-dimensional marginal slices. Appropriate software for these kinds of applications is being developed (see Tierney, 1991).

Consequentially, computing power permits more sophisticated interpretations of phenomena. No longer should a researcher be tempted to do a simple χ^2 on each of four separate 5 x 7 tables. Rather, the researcher can treat the full 4 x 5 x 7 table, using log-linear models (see, e.g., Bishop, Fienberg, & Holland, 1975; Fienberg, 1977). Log-linear models express the probability of an occurrence (say, falling in a certain cell of a contingency table) as a linear function of specified variables (say, the three marginal categories of the 4 x 5 x 7 table and their interactions). Although generally involving maximum likelihood procedures that are computationally intensive, analysis based on log-linear models is readily available with standard statistical software such as SAS and SPSSX. The computational excuse gone, the researcher must face the challenge of dealing with the interpretation of discovered higher-order interaction effects. Switching to a related field, without the computational excuse of the ease of linear regression, the researcher is compelled to examine the perhaps more conceptually valid class of models that require nonlinear regression (see, e.g., Draper and Smith, 1981).

The thirst for more computing power will continue, nevertheless. Today at Carnegie Mellon University, for example, researchers are examining the use of neural networks (see, e.g., Hertz, Krough, & Palmer, 1991) for nonlinear discriminant analysis in predicting criminal recidivism. A widely used neural network approach is the multilayer perceptron. The basic building block of the multilayer perceptron is the single perceptron, which plays the role of an artificial neuron. The perceptron compares a weighted combination of inputs—say number of prior arrests, age of the prisoner and severity of offense—to a threshold value and produces a binary output. The weights are determined using supervised learning, that is, by presenting a number of inputs with known outputs and adapting the weights accordingly. The single perceptron essentially provides a linear discriminant function and can only solve problems that are linearly separable in the input space. This restriction is relaxed by connecting perceptrons in layers and using a nonlinear transfer function between layers. Training data are again used to determine the weights of the neural network with a back-propagation algorithm being used to minimize the discrepancies between the observed and predicted network outputs. Once a network has been trained it can be used to classify new data. This prediction device corresponds to a non-linear discriminant function. Building a neural network is computationally intensive. One study applied this procedure to R. A. Fisher's time-honored iris data (Fisher, 1936) in which the problem was to discriminate between two varieties, Iris Versicolor and Iris Setosa, on the basis of multivariate measurements of sepal length, sepal width, petal length, and petal width. With 50 observations on each species the standard discriminant analysis computation would take seconds on most personal computers; but the neural network computation took more than 40 minutes on a SUN workstation. We can fully expect that methodological developments will continue to challenge the ever expanding computational frontier.

Computer Networks

Computer networks, such as BITNET, already link researchers worldwide. These networks are becoming more standardized and comprehensive, and are linked together in a world-wide super network of perhaps 25 million users called Internet (Krol, 1992). With this expanded capability of computer networks (and further links to such popular networks as CompuServe), a researcher at Ohio State, say, can cheaply obtain data from a plant in Belgium and exchange ideas, papers, and data with a researcher at Seoul National University. Collaborations will be easier, and physical distance will be much less constraining. These networks will also permit researchers to access large databases quickly. Pushing for further development of computer network capability, The Electronic Frontier Foundation (headed by Mitch Kapor, the founder of Lotus Development Corporation) is a group "that seeks to develop public policies to maximize the social potential of new computer and communications technologies." Although the technology will be there, a serious constraint on its use will be privacy and confidentiality concerns, as the most valuable databases will contain sensitive microdata on individuals, households and organizations.

Data Capture and Experimentation

Most often in management research, data capture is either through observational studies or through designed experiments. Cochran (1983) notes that observational studies have two characteristics:

1. the objective is to study the causal effects of certain agents, procedures, treatments or programs.

2. for one reason or another, the investigator cannot use controlled experimentation, that is, the investigator cannot impose on a subject, or withhold from the subject, a procedure or treatment whose effects he desires to discover, or cannot assign subjects at random to different procedures (p. 1).

Data Capture and Causal Inferences

Both observational studies and designed experiments, then, deal with data capture for inference about causal relationships. Where they differ is in the researcher's capability to determine the assignment of treatments and subjects. Typically, observational studies have the advantage that they capture data on real management practice in real organizations. Their disadvantage is the lack of researcher control on assignment of interventions to subjects. This makes it tenuous to infer causal effects from the observational data. As Roberts (1991) notes, "The real question is whether really useful causal information can be made from the kinds of happenstance data typically available in organizations" (p. 15). I am pessimistic about an affirmative answer, agreeing with Barnard (1982) and others that if you want to know what happens when you change something, you must change it and see.

Data Capture and Continuous Quality Improvement

To better address causal inference, an exciting possibility is to couple new statistical and computing capabilities to enable experimentation and evaluation analysis as a normal component of management practice. Specifically, online databases can capture data in the usual operation of an organization and these data can be analyzed in real time as a decision support system to management. Planned departures from standard operating procedure can be introduced to provide systematic information leading to continual improvement of the process. If these planned departures are small and if this experimental program is taken to be part of a long-term search for improvement in management practice, these innovations need not materially disrupt the flow of work in the organization. This idea is not new, having been suggested by Box and Draper (1969) for industrial processes under the heading of EVOP, standing for Evolutionary Operations. What I suggest is that much of the technology is here now for its broader application to management in organizations, ranging from scheduling the staffing of a fire department to continuous improvement of surgical procedures in clinical use. In many cases it requires knowledge of experimental design and systematic use of multivariate time series methodology. The emphasis is on a dedication to continual quality improvement, forever seeking out both sporadic and systemic impediments to quality. Implementing this approach can lead to much closer and continued involvement between the management researcher and the organization under study.

Privacy, Confidentiality, and Data Access Concerns

Increasingly, management has the capacity to monitor, store and analyze in real-time every keystroke of a data entry clerk, every word of a telemarketer, and every decision of a 911 dispatcher. Extending beyond the organization's own employees, management can also examine the complete purchasing history of an individual supermarket customer, the response of potential donors to political contribution appeals, and the pattern of submission of false alarms at emergency call boxes. This capacity for data capture and analysis both expands possibilities for effective management control and provides new opportunities for management research, as we have argued in the previous section. In particular, just as large data bases like CRSP and COMPUSTAT opened up possibilities for research studies in finance at a fairly macro level, these rich data bases for individual plants, offices and stores open up possibilities for the kind of micro level research studies that can lead to real insight into management practice. On the other hand, it also raises important ethical and pragmatic issues of privacy and fairness (Duncan & Pearson, 1991).

This concern for privacy and fairness is accentuated when the organization finds it profitable to sell its accumulating database to other interested organizations, as some credit card companies sell purchasing databases and many state motor vehicle departments sell drivers' license data and traffic citation history. The issue of providing security to statistical databases against disclosure of confidential information is a problem of practical concern to database administrators. A single database can serve both administrative and statistical functions. For example, a medical database is used by physicians to clinically monitor individual patients; this is an administrative function since information supplied is relevant to the treatment of a particular patient. On the other hand, public health researchers require only aggregate statistics, since their goal is population inference. A statistical database can be thought of as a shell for a database in which authorized users are entitled to access such quantities as counts, averages and regression coefficient estimates. They are not entitled to obtain, directly or through inference, the identity of an entity. In typical applications, where sensitive information, such as income, criminal history, and sexual practices may be stored, such identification would be a breach of confidentiality.

Inferential Disclosure

Direct identification is easily controlled by database software that refuses access to fields containing individual identifiers such as name and social security number. However, even when such identifiers are suppressed, it may be possible to infer confidential information about individuals based on available information. This type of disclosure is called inferential disclosure, and is much more difficult to control. Work on this topic is proceeding (see, e.g., Keller-McNulty & Unger, 1991; Duncan & Mukherjee, 1991).

Because of legitimate concerns about the possibility of disclosure of individual information, government statistical agencies have limited the amount of detailed data provided to nongovernment users in tabulations and public-use microdata files. This lack of detail limits the ability of users to do analyses that could contribute to the understanding and resolution of significant economic and social problems. Some agencies have developed mechanisms for providing access to more detailed information on a restricted basis, but present arrangements do not meet all legitimate needs.

Expanding Data Access

Statistical agencies should expand their efforts to make more detailed data available to users, subject to appropriate restrictions to ensure confidentiality. Licensing of data users, as recently developed by the National Center for Education Statistics, is a promising approach. Users, for their part, should be aware of the importance of maintaining confidentiality of individual information and should do everything within their power to prevent disclosure. I believe that those allowed restricted access to data sets should be subject to legal sanctions for failure to comply with agreed-on conditions of use.

Privacy Erosion

The public feels increasingly, and with some justification, that its privacy is being eroded by organizations that develop and control the use of large databases that contain detailed personal information. Linkage of data from different sources is perceived as a particular threat. For these and other reasons, statistical agencies, whether public or private, are finding it more difficult to persuade persons and organizations to participate in statistical surveys, whether voluntary or mandatory. Statistical agencies should take several steps to meet the privacy concerns of the public and to promote continued public support of vital data collection programs. These include:

1. Ensuring "informed consent"—that all survey respondents are given clear explanations of the conditions and consequences of their participation.
2. Undertaking and supporting research, using the tools of cognitive and survey research, to monitor views of the public on informed consent, response burden, sensitivity of survey questions, data sharing for statistical purposes and related issues.
3. Actively soliciting the views of advocacy groups that are concerned with privacy issues. Where appropriate, representatives of these groups should be included on advisory committees.
4. Exploring the possibility of greater use of administrative record sources to meet some information needs.

Causation

Empirical research, in general, has purposes of prediction, description, understanding and control. Each of these has its place in management research. Prediction is essential to the manager who must act now—say to employ workers, borrow money and build plants—in expectation of uncertain future circumstances—like product demand, price levels and inflation levels—hence there is an appropriate emphasis on forecasting as a management tool. Statistical description helps communicate characteristics of aggregates, as a discussion based on median salary levels may ease decisions about appropriate executive compensation. Understanding of complex processes, like highway contract bidding, may further the development of more advantageous procedures.

At its heart, though, management research is a search for control mechanisms, for causal ties. Managers want to know what they can change that will produce better outcomes. Here even the capitalist-oriented might agree with Karl Marx, "The philosophers have only interpreted the world in various ways; the point, however, is to change it."

Conditions for Causality

Generally a relationship is taken as causal if three conditions are satisfied (adapted from Babbie, 1992):

1. The cause precedes the effect in time.
2. There is an empirical correlation between the cause and the effect.
3. This correlation is not the result of the effects of some third variable—otherwise the observed correlation would be spurious evidence of causation.

Causality is an area that has been and is being intensely investigated. Clearly, the hard part is establishing the third condition—ruling out the effects of other variables or rival explanations. A useful and nearly classical approach to causality is that of Paul Lazersfeld, called "elaboration" and based on the idea of conditional independence. The idea is that X causes Y if for any variable Z, the conditional correlation of Y and Z given X is zero. Let's see how elaboration works with the notion that if a woman supports abortion rights, this causes her to be favorably inclined toward political candidates that are Pro–Choice. Given a woman's position on abortion there should be no correlation between support of large military expenditures and support of a Pro–Choice candidate. As this example suggests, while "elaboration" is elegant it is excessively strong in that it does not admit multiple supportive causes. Other approaches that are employed include effects analysis, due first to Finney (1972) and importantly to Alwin and Hauser (1975), and path analysis due to Wright (1934) and developed in a social science context by Blalock (1961). Some promising work is now being done to discover causal structure in the case of simultaneous linear equation models (See Pearl, 1988; Spirtes, Glymour, & Scheines, 1991; Spirtes, Scheines, & Glymour, 1990). Causality is a hard problem that deserves considerable attention by all researchers—and certainly management researchers.

Modeling and Simulation

Whether the purpose of management research be prediction, description, understanding or control, a statistical model allows a complex reality to be workably expressed. Such a model is generally clearly defined in mathematical terms and so is relatively easy to communicate to anyone familiar with mathematical notation. This makes it possible to examine readily its strengths and weaknesses, taking advantage of mathematical manipulation or computer simulation (Gilchrist, 1984).

In this modeling area, I expect continued developments in three areas: (1) greater use of computing power to model more complex interactions, (2) less use of model-independent statistical hypothesis testing, and (3) easier to use computer software for modeling and simulation.

Modeling Complex Interactions

With increased computing capability, including in the computer graphics area, modeling and simulation methods will become increasingly sophisticated and allow a more detailed exploration and evaluation of a greater range of management mechanisms. This will allow context to be more fully explored. I will illustrate the current problem and suggest some potential benefits of this in the forecasting area. Chelimsky (1991) notes that "the Bureau of Labor Statistics 1970 forecast of the 1980 U.S. economy entirely missed the massive entry of women into the labor force that was to occur between 1970 and 1980" (p. 228) and asserts that such events are unpredictable. Yet, the technological and social changes that made this entry of women into the labor force were well in place by 1970; in particular the economic advantage of muscle power was lost to brain power and the baby boom that limited the labor market options of women was over. More detailed modeling of this technological and social context could well have brought out this potential for women to enter the labor force.

Methodologically, new tools are being developed to address context. For example, Duncan, Gorr, and Szczypula (1993) developed forecasting methods that draw on a multitude of related time series, each modeled as a multi-state Kalman filter. A case study of forecasting income tax revenue for each of forty school districts in Allegheny County, Pennsylvania, based on fifteen years of data, is used to illustrate the procedure. Originally applied in engineering, the Kalman filter is a recursive, unbiased least squares estimator of a Gaussian stochastic process (see, e.g., Jazwinski, 1970). The Kalman filter is now also being used in the biological and social sciences (see, e.g., Abraham and Ledolter, 1983)). The multi-state Kalman filter is particularly suited for short and irregular time series data where structural changes of a transient, step, or trend nature occur. Structural change is a prevalent problem in local government revenue forecasting, where national or local economic booms and busts often dramatically impact tax collections. It is also a common problem in sales forecasting, because of promotional activities and competitor's actions (Lewandowski, 1982; Makridakis & Wheelwright, 1989). For planning and operational management purposes, many organizations forecast over three dimensions: (1) *products* (consumer products, public services, revenue generating activities such as collecting taxes, etc.), (2) *locations* (sales territories, municipalities, etc.), and (3) *time* (months, quarters, years, etc.). A multivariate time series model would seem to be especially appropriate for forecasting over products, say PC sales and spreadsheet sales, as these values are intrinsically correlated (Tiao & Box, 1981). On the other hand, the multiple time series generated by total sales for various computer store locations may have a correlational structure that is generated only by dependence on outside factors such as the state of the economy or trends in computer usage. Also the number of computer store locations may be very large so that forecasting using direct multivariate approaches would require the specification of an excessively large number of parameters to capture the correlational structure. The approach used by Duncan, Gorr, and Szczypula (1993) makes uses of a hierarchical model in which the probability distribution for the cross-sectional measurements has the same form for each observational unit, but the parameters of that distribution vary over observational units. This provides a mechanism for building context into the model.

A Decline of Hypothesis Testing

Model-independent statistical analyses aimed simply at confirming statistical significance will become less important, as researchers seek to better specify the nature of relationships. Thus hypothesis testing, with its dichotomous, "yes, significant vs no, not significant", conclusion will not be the primary statistical tool used in management research that it seems to have been in the past. Instead more emphasis will be placed on modeling complex structure based on large data sets. The art will be in developing relatively simple representations of the managerial phenomenon, even though conventional hypothesis tests would reject such simple models because of the large sample size. Further, interest is growing in meta-analysis of information derived from multiple sources. For example, from 1970 to 1979 aspirin was compared with a placebo based on mortality rate after heart attack in six multicenter randomized clinical trials (Canner, 1987). Conventional hypothesis tests yielded one-sided observed significance levels (P–values) favoring aspirin that ranged from .028 to .898. Simply aggregating the data from the six trials yielded a P–value of .072. For a physician managing patient care, what should be made of these results? As demonstrated in Gaver et al. (1993), simple combinations of P–values, say by Fisher's composite P–value (calculated by treating the sum of twice the negative logarithm of k individual P–values as a χ^2 random variable with 2k degrees of freedom), have undesirable properties. An appropriate answer to the physician's questions would seem to only come from building a hierarchical model.

Easier to Use Software

Modeling and simulation software will become easier to implement and more flexible in dealing with management processes. The proliferation of personal computers will make it economically feasible for software firms to develop easy-to-use packages that will provide the researcher with workable tools. Minitab, a software system that is widely used in introductory statistics courses, has over the years increased in power and ease of use. A version for Microsoft Windows has recently been introduced.

As an example of software to address more complex statistical modeling, LISREL (LInear Structural RELationships) has been found useful in developing structural equation models throughout the social and behavioral sciences. Originated by Jöreskog (1973), LISREL covers factor analysis models, path analysis models, models for time-series data, recursive and non-recursive models for cross-sectional and longitudinal data, and covariance structure models. An important feature of the LISREL modeling approach is that the variables in the equation system can be either directly observed or unmeasured (latent). It can handle problems involving measurement errors and interdependent causation. Estimation can be through the method of least squares, instrumental variables, or maximum likelihood.

The capacity to do the computation to fit models will grow. For example, maximum likelihood estimates for complex models can now be obtained using a variety of software packages for the PC, such as Gauss. Essentially a high-level programming language, Gauss treats every variable as a matrix and has the capability of solving systems of nonlinear equations. Although Gauss is somewhat difficult to learn (MacKie-Mason, 1992), it can be expected that future incarnations will be increasingly user friendly.

The Bayesian Perspective

The Bayesian perspective on statistical inference and decision analysis calls for all unknown quantities to have uncertainty about their values modeled by a joint probability distribution. For years it has been accepted that the Bayesian perspective has certain theoretical advantages. First, it provides a conceptually simple three-stage algorithm for solving any inference problem (Box & Tiao, 1973). Stage 1 specifies the prior distribution: the joint probability distribution of all unknown quantities. Stage 2 specifies the form of the conditional probability distribution of the data given the unknown quantities. Stage 3 uses Bayes' rule to compute the posterior distribution: the joint probability distribution of all unknown quantities given the observed data values. Second, the Bayesian perspective is the only one that satisfies certain appealing axioms of coherence in inference (De Groot, 1970). Until recently, the application of Bayesian statistics has been limited. As Lindley (1982) notes about the Bayesian method:

> Its potential is enormous, but its implementation requires extensive calculations, principally in the performance of numerical integration (to remove nuisance parameters and to find expectations) and maximization (to select optimal decisions). Suitable computer packages to do these calculations, and to interrogate the scientist on his or her knowledge about the situation and convert it into probability forms, have just been written. Until they are more widely available, applications to elaborate sets of data are bound to be limited (p. 204).

Recently, statistical researchers have shown that with increased computing power and sophisticated numerical methods the Bayesian perspective can be used to deal systematically and reasonably with a wide range of practical statistics problems. The success over the last decade of the semi-annual Seminar on

Bayesian Inference in Econometrics and Statistics organized by Arnold Zellner of the University of Chicago in encouraging applied research is evidence of this. Further evidence that Bayesian statistics is being taken out of the realm of the purely theoretical is the Practical Bayesian Statistics Conference held at the University of Nottingham, July 8–11, 1992 and the Bayesian Robustness International Workshop held in Milan, May 18–21, 1992. Particularly interesting is the development by Bruce Hill of the University of Michigan of a Bayesian data analysis as an integration of the informal tools of data analysis, as popularized by John Tukey with Exploratory Data Analysis (EDA) (see Tukey, 1977), with the concepts and methods of subjective probability.

In the past a major drawback to the application of Bayesian techniques has been the lack of computational capability. Even the most ardent Bayesian would in practice be led to maximum likelihood estimates and asymptotic theory. However, developments in statistical computing, especially in multivariate integration, have overcome some of the technical difficulties in applying Bayesian inference tools. Through the use, say, of Laplace's method for the asymptotic expansion of integrals, effective computer techniques can be constructed for approximating posterior expectations.

Moving to the more psychological arena, cognitive research has made progress in identifying the mechanisms whereby people deal with uncertainty, and this has implications for the assessment of prior distributions. In the psychology literature, some useful references are Kahneman, Slovic, and Tversky (1982), and Hogarth (1987). In the decision analysis area, a useful reference is von Winterfeldt and Edwards (1986). In the statistics area, see Chaloner and Duncan (1983, 1987), Garthwaite and Dickey (1988), and Chaloner, Church, Louis, and Matts (1992).

The Bayesian perspective requires that the prior assessment be explicitly produced. This may have the salutary effect of making public the personal viewing lenses through which researchers and managers view problems. In this regard, Vincent Barabba, former Director of the Census Bureau and now Executive Director of Market Research and Planning for General Motors, remarks in his 1990 Presidential Address to the American Statistical Association:

> The "Law of the Lens" . . . manifests itself in many ways. In business, for example, new product management teams, when assessing the feasibility of introducing new products, tend to favor their introduction with a frequency that is unwarranted by subsequent commercial performance. Analysis of many new product failures points to a tendency to both design research, and interpret research results, in a way that does not allow equal opportunity for evidence to arise that runs counter to a new product launch decision. Based on my experience in government, I could say the same of agency heads attempting to support their programs in either Office of Management and Budget analysis or congressional budget committees.

Discussion

Progress, especially in computer and communications technology, has indeed raised many problems for the management researcher—but also many opportunities. These opportunities allow the researcher to get closer to the operative processes in an organization, to apply more effective techniques and to draw better causal inferences. Today, for the researcher to better understand and advance management practice requires a methodological command of computer and communications capability, a knowledge of sophisticated data capture procedures, an understanding of the principles and techniques of social science experimentation, an appreciation for the need to protect confidentiality and to share data under appropriate circumstances, an understanding of how to assess causal relations, a knowledge of modeling methods and simulation techniques, and an understanding of Bayesian statistics. These issues have been addressed in the article, but there are other issues that the management researcher may have to address. For one example, there is much public discussion that the nature of management is changing. Much of what is being discussed suggests a greater participatory role for workers at every level of the organization. This is one credo of Total Quality Management (TQM), for instance. If these changes are widespread, they can affect the nature of management research methodology, perhaps through greater emphasis on carefully-composed focus groups. For a second example, alternative dispute resolution (ADR) approaches such as mediation and arbitration have for years been used to deal with labor/management disputes. Lately, however, there has been an upsurge of ADR application in areas that border on management practice, such as small claims mediation and "reg neg" or regulation negotiation. These and like ADR applications could provide fertile ground for management research. Methodologically, the growth of these applications may, as with the first example, lead the management researcher to explore innovative data capture procedures.

References

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting.* New York: Wiley.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47.

Babbie, E. (1992). *The practice of social research.* Belmont, CA: Wadsworth.

Barnard, G. A. (1982). Causation. In Kotz, S., & Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, pp. 387–389). New York: Wiley.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Blalock, H. M., Jr. (1961). *Causal inferences in nonexperimental research.* Chapel Hill, NC: University of North Carolina Press.

Box, G. E. P., & Draper, N. R. (1969). *Evolutionary operation.* New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* Reading, MA: Addison-Wesley.

Canner, P. L. (1987). Aspirin in coronary heart disease: Comparison of six clinical trials. *Statistics in Medicine, 6,* 255–263.

Card, J. J., & Peterson, J. L. (1991). Establishing and operating a social science data archive. In Sieber, J. E. (Ed.), *Sharing social science data: Advantages and challenges* (pp. 116–127). Newbury Park, CA: Sage.

Chaloner, K., Church, T., Louis, T., & Matts, J. (1992). *Graphical elicitation of a prior distribution for a clinical trial.* (Technical Report No. 583). St. Paul, MN: School of Statistics, University of Minnesota.

Chaloner, K., & Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *Statistician: Journal of the Institute of Statisticians, 32,* 174–180.

Chaloner K., & Duncan, G. T. (1987). Some properties of the Dirichlet–multinomial distribution and its use in prior elicitation. *Communications in Statistics: Theory and Methods, 16,* 511–523.

Chambers, John M. (1977). *Computational methods for data analysis.* Wiley: New York.

Chelimsky, E. (1991). On the social science contribution to governmental decision-making. *Science, 254,* 226–230.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74,* 829–836.

Cochran, W. G. (1983). *Planning and analysis of observational studies.* New York: Wiley.

Date, C. J. (1983). *Database: A primer.* Reading, MA: Addison-Wesley.

David, M. H. (1989). Managing panel data for scientific analyses: The roles of relational data base management systems. In D. Kasprzyk, G. J. Duncan, G. Kalton & M. P. Singh (Eds.), *Panel surveys* (pp. 226–241). New York: Wiley.

David, M. H. (1991). The science of data sharing: Documentation. In J. E. Sieber (Ed.), *Sharing social science data: Advantages and challenges* (pp. 91–115). Newbury Park, CA: Sage.

De Groot, M. H. (1970). *Optimal statistical decisions.* New York: McGraw-Hill.

DiCiccio, T., & Romano, J. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B, 50,* 338–354.

Draper, N., & Smith, H. (1981). *Applied regression analysis,* (2nd Ed.). New York: Wiley.

Duncan, G. T., Gorr, W., & Szczypula, J. (1993). Bayesian forecasting for seemingly unrelated time series: Application to local government revenue forecasting. *Management Science, 39,* 275–293.

Duncan, G. T., & Mukherjee, S. (1991). Microdata disclosure limitation in statistical databases: Query size and random sample query control. *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy,* 278–287.

Duncan, G. T., & Pearson, R. W. (1991). Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Statistical Science, 6,* 219–239.

Efron, B., & Tibshirani, R. (1991). Statistical data analysis in the computer age. *Science, 253,* 390–395.

Federer, W. T. (1973). *Statistics and society: Data collection and interpretation.* New York: Dekker.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge, MA: MIT Press.

Finney, J. M. (1972). Indirect effects in path analysis. *Sociological Methods and Research, 1,* 175–186.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7,* 179–188.

Garthwaite, P. H., & Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B, 50,* 462–474.

Gates, W. A. (1989). A relational database structure for the March Current Population Survey. *Proceedings of the Bureau of the Census annual research conference* (pp. 59–68). Washington, DC: U.S. Department of Commerce.

Gaver, D. P., Draper, D., Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., & Waternaux, C. (1993). *On combining information: Statistical issues and opportunities for research panel on statistical issues and opportunities for research in the combination of information.* Washington, DC: National Research Council.

Gilchrist, W. (1984). *Statistical modelling.* New York: Wiley.

Hertz, J., Krough, A., & Palmer, R. G. (1991). *An introduction to the theory of neural computation.* Reading, MA: Addison-Wesley.

Hogarth, R. (1987). *Judgement and choice* (2nd ed.). New York: Wiley.

Hurley, C., & Buja, A. (1990). Analyzing high dimensional data with motion graphics. *SIAM Journal of Scientific and Statistical Computing, 11,* 1193–1211.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory.* New York: Academic Press.

Jöreskog, K. G. (1973). In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Seminar Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgement under uncertainty: Heuristics and biases.* Cambridge, MA: Cambridge University Press.

Keller-McNulty, S., & Unger, E. A. (1991, March). *Database systems: Inferential security.* Conference on Disclosure Limitation Approaches and Data Access of the Panel on Confidentiality and Data Access. Washington, DC: National Academy of Sciences.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic.

Krol, E. (1992). *The whole internet users guide and catalog.* Sebastapol, CA: O'Reilly & Associates.

Lewandowski, R. (1982). Sales forecasting by FORSYS. *Journal of Forecasting, 1,* 205–214.

Lindley, D. V. (1982). Bayesian inference. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, pp. 197–204). New York: Wiley.

MacKie-Mason, J. K. (1992). Econometric software: A user's view. *Journal of Economic Perspectives, 6,* 165–187.

Makridakis, S., & Wainwright, S. (1989). *Forecasting methods for managers* (5th Ed.). New York: Wiley.

Miller, J. J., & Wegman, E. J. (1989). Contraction of line densities for parallel coordinate plots. In K. Berk & L. Malone (Eds.), *Computing science and statistics: Proceedings of the 21st symposium on the interface* (pp. 191–199). Alexandria, VA: American Statistical Association.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Roberts, F. S. (1979). *Measurement theory.* Reading, MA: Addison-Wesley.

Roberts, H. V. (1991, April). *Total quality management: A research agenda.* Paper presented at the 1991 AACSB annual meeting, St. Louis, MO.

Samiuddin, G. M., & El-Sayyad, G. M. (1991). On nonparametric kernel density estimates. *Biometrika, 77,* 865–874.

Spirtes, P., Scheines, R., & Glymour, C. (1990). Simulation studies of the reliability of computer-aided model specification using the TETRAD II, EQS, and LISREL programs. *Sociological Methods and Research, 19,* 3–66.

Spirtes, P., Glymour, C., & Scheines, R. (1991). From probability to causality. *Philosophical Studies, 64,* 1–36.

Tiao, G. C., & Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association, 76,* 802–816.

Tierney, L. (1991). *LISP–STAT: An object-oriented environment for statistical computing and dynamic graphics.* New York: Wiley.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge, MA: Cambridge University Press.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.

Biographical Note

George T. Duncan
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15217
Phone/Fax: 412–268–2172/7036

Dr. Duncan is Professor of Statistics, Carnegie Mellon University. Duncan's current research interests focus on privacy and information issues. He chairs the National Academy of Sciences Panel on Confidentiality and Data Access. He has been the editor of the *Journal of the American Statistical Association.* His research work has been published in such journals as *Biometrika, Journal of Conflict Resolution, Journal of the American Statistical Association, Management Science, Operations Research* and *Statistical Science.*

Received: September 15, 1992
Accepted after two revisions: July 10, 1993

♥