R-1375
April 1975

Methodology for Subjective Assessment of Technological Advancement

S. James Press and Alvin J. Harman
With Appendix by Marc Nerlove and S. James Press

Rand
SANTA MONICA, CA. 90406

R-1375
April 1975

Methodology for Subjective Assessment of Technological Advancement

S. James Press and Alvin J. Harman
With Appendix by Marc Nerlove and S. James Press

Rand
SANTA MONICA, CA. 90406

Published by The Rand Corporation

PREFACE

This report addresses four questions pertaining to the allocation of funds for research and development:

1. How can we determine future military deficiencies for which technological advancements may provide remedies?
2. What tools can be developed to relate these deficiencies to the range of alternative technological remedies?
3. What procedures can be used to assess and compare the various remedies for their potential for technological advancement?
4. How could the information developed in (1), (2), and (3) be used for management purposes?

Various research strategies are identified in pursuit of answers to these questions. In the context of the third question, analytical techniques are discussed which appear promising for quantifying technological advancement.

The mathematical Appendix to this report first appeared in Rand Report R-1306-EDA/NIH, Univariate and Multivariate Log-Linear and Logistic Models, by Marc Nerlove and S. James Press, December 1973.

This discussion of the issues is intended as a guide to approaches to program planning by Department of Defense groups undertaking or funding exploratory R&D.

SUMMARY

This report is concerned with providing a policymaker with information that will help him in his choice of exploratory R&D options--options that will subsequently affect the feasibility or availability of new systems. The fundamental issue is, How do we assess the relative degree of technological advancement of different types of projects (such as those associated with enhancements of aircraft and missile capabilities) that may require long periods of research and development, if they are feasible at all?

In this study we approach the problem by first relating the assessment of technological advancement to R&D program management objectives. Broadly defined, program management, as practiced by the major federal R&D funding agencies, consists of four interacting activities: planning, development, evaluation, and utilization. Using these activities as a framework, it is possible to determine areas in which the assessment of technological advancement may make a difference, the degree depending on whether we are considering a long-term or short-term horizon. In long-term program planning, for instance, assessment of technological advancement can be very useful in gathering information on a technology or program objective in order to identify important new prospects for R&D, and to indicate what resources will be required to pursue them.

Since technological-advance assessments are obviously important to program planning and development, how can such assessment be obtained? The first step is to determine the scope of the assessment problem by identifying which potential advances will be useful and then considering different technologies that may be alternative solutions to the same problems. Once we have established what our assessment objectives are, we can develop criteria for deciding who is qualified to make appraisals of the potential advances in those objectives, and then design procedures for obtaining and grouping expert judgments.

Basic to our approach to technological-advance assessment is the notion that expertise does exist, that there are people who have considerable knowledge and understanding of the mechanisms underlying

particular problems and thus can do an appreciably better job of forecasting long-term trends and changes than the non-expert. But how do we recognize an expert? How many experts constitute a "good" panel? What approach should be used in eliciting their responses? What types of questions should be asked, and how should they be structured? This study addresses these and other important issues associated with resolving the procedural and administrative problems involved in selecting a panel of experts, in eliciting informed judgments about the degree of technological advance or relevant projects, and in designing a survey questionnaire for measuring those judgments.

Three methods of multivariate analysis are described for quantifying and analyzing group judgment data collected from a panel of experts. Those of the methods that are known in earlier literature include multidimensional scaling of individual differences, and subjective probability procedures, including the Bayesian approach. A new procedure, which was developed specifically for this application, involves the use of simultaneous equation system models in which the response (dependent) variables are categorical, i.e., discrete (possibly having a value of zero or one, only) and unordered. Such dependent variables are interpretable in terms of probabilities. Since project feasibility can be measured by the probability that a new capability can be developed, and this probability can be related to other probabilities that various components can be developed, the new procedure described in the report has great potential applicability. (The details of this new statistical procedure are provided in an appendix.)

Further research is still needed in a number of areas. For example: (1) The methods for selecting a panel of experts should be refined. (2) Planning objectives should be devised for selecting technology areas. (3) Criteria should be established for determining the number and type of experts. (4) An interrogation procedure should be developed, with the questions designed to assess technological advancement in specific technologies. Also, at least two types of analysis should be undertaken: individual-differences scaling using a multistage controlled feedback approach, and a categorical-dependent-variable multivariate regression (but without the feedback data).

In short, the material presented in this report is as much proposal for the initiation of a technological-advance assessment based on the principles and procedures described herein as it is a presentation of research findings.

ACKNOWLEDGMENTS

A number of colleagues have contributed ideas and suggestions to this work. They are H. Bailey, A. Brewer, N. Feldman, S. Fienberg, L. Goodman, W. B. Graham, H. Henning, A. Lieberman, C. N. Morris, L. Mundie, D. Relles, T. P. Schultz, G. K. Smith, J. P. White and C. Wolf, Jr. We are grateful to Darien Roseen, John Rolph, Bridger Mitchell, and Dorothy Stewart for editorial comments on various portions of the manuscript. None of our colleagues should be held responsible for deficiencies in this report.

CONTENTS

I. INTRODUCTION

This report presents techniques and approaches for assessing the feasibility of new systems or projects. We discuss ways of generating information that will help us to better articulate a "menu" of R&D options, and ways of developing analytical tools for refining the comparison among options. We also show how technical and economic expertise can be used: for assessing technological advances and trends, using quantitative techniques; for comparing technological advances for various projects; and for determining how technological advances affect R&D costs.

Because there are few inevitable or unalterable technological trends, and few fixed intervals between experimental and operational hardware, it is quite important that we improve our capability for assessing technological-advance prospects in order to improve planning and management activities for the more heterogeneous R&D projects. That is, managers should *influence* technological advance through deliberate action.

In this report, we propose a methodology that will help to determine the allocation of R&D and thus improve the *rates* of technological advance and *reduce* the feasible *leadtime* from experimental to operational capabilities. To achieve these goals, we must have a thorough understanding of (1) the "value" of a successful technological advance of any objective, (2) the scope of the objectives that determine which technologies should and should not be considered as alternatives, (3) the procedures for assessing and comparing technological-advance alternatives, and (4) the management activities into which the technological-advance assessment results will fit. Here we focus most heavily on topics two and three--the use of broad objectives in developing methods for acquiring and analyzing information with which to compare the feasibility of various technological advances. Topics one and four must be considered, however, if the end results of the assessment are to be useful.

Past research at Rand has yielded methodologies for assessments

of technological advance that concentrated on "outputs" of R&D leading to improve military capabilities (as a measure of technological "success"), rather than on specific technologies (as "inputs" to improve capability) that may or may not be economically efficient to pursue.[1] For example, Rand's turbine engine research concentrated on an assessment of technological trends based on performance advancement rather than on the specific considerations of compressor designs or use of advanced alloys. Maintaining such trends may call for different "input-technology" solutions at different times, and trying to "push" ahead of the trend may become very costly or risky. The resulting trend line could be characterized as one of technology embodied in operational hardware. Conceptually, one can imagine such trends as extending from the advancement of scientific knowledge (as it applies to improved military capability) to the varying degrees of embodiment in experimental breadboard or brassboard, prototype, or production hardware. Although the analytical methods of technological-advance assessment thus far developed are appropriate to R&D projects that lead to equipment with the same set of operational attributes over a long period of time, these methods are not directly appropriate to more heterogeneous R&D projects and component advancements.

To develop sound methods for such applications, we have looked for the best way of using historical data and subjective expertise in assessing the advancement attainable in future projects. If we use a panel of "experts" in relevant fields, there are a number of important factors that must be considered. We must first determine who should participate in an exercise to evaluate alternative technologies, and then devise procedures to formulate carefully a broad range of questions with respect to technologies, resource requirements, and time horizons that would be required to make relative comparisons

---

[1] Previous work in these general areas include A. J. Alexander and J. R. Nelson, *Measuring Technological Change: Aircraft Turbine Engines*, R-1017-ARPA/PR, June 1972; A. J. Harman, *Choice Among Strategies for System Acquisition*, The Rand Corporation, P-4794, March 1972; J. R. Nelson and F. S. Timson, *Relating Technology to Acquisition Cost: Aircraft Turbine Engines*, The Rand Corporation, R-1288-PR, March 1974.

across a technological area.[2] We must also give serious consideration to the kinds of incentives that should be provided to the participants, so that they will take the exercise seriously and provide thoughtful, conscientious responses based on their expertise. Procedures for eliciting expert judgments are discussed in Section III. For example, one way of eliciting information would be to request it in terms of alternative "paths" to the planning objective, at the same time obtaining the probabilities of success of each node in the path and the plausible range of time and resource commitments needed for a reasonable probability of success. Such information would provide valuable insights into promising technological areas for R&D support.

Once we have elicited information from a panel of experts, we will need valid, appropriate methods of analyzing this information. The raw information may be of direct interest to the extent that detailed responses are obtained from some of the experts. It may be desirable to make careful engineering evaluations of the responses of the group and to use in-house judgments to provide decision alternatives. But it also seems quite likely that insights can be obtained by using more formal statistical procedures for assembling the group judgments. Such procedures are proposed in Section IV.

Before we can discuss technological assessment techniques in detail, however, we need to specify which phase of the planning or operation of an R&D program we are addressing. *We assume that we are at a very early phase of planning a new set of research projects.* Therefore, techniques of analysis should yield a broader but less thorough understanding of individual project areas than those for project selection or evaluation. This is the subject of Section II.

We are attempting to build on our past experience in developing techniques for assessing technological advance. Such assessments may provide important information for decisionmaking, especially if we consider the prospects for advance in the context of the resources needed

---

[2]The kinds of persons to be involved in any one technological assessment will clearly depend on the scope of the technologies involved; but in most cases a broad range of perspectives, from both industrial and university backgrounds, may be desirable.

to support the effort. But other information is desirable: Is the subject of R&D being ignored completely by other supporters (in the DoD and other federal agencies) and by private industry? What are our allies doing in the area? The Soviets? What are their future plans? Thus the assessments and related policy issues need to be addressed in terms of specific technologies, and it is this aspect of the decision process on which we will focus.

II. PROBLEM DEFINITION: MANAGEMENT OBJECTIVES IN MEASURING TECHNOLOGICAL ADVANCEMENT

MANAGEMENT FRAMEWORKS[3]

We can rather broadly characterize a management function as consisting of four major *interacting* parts: (1) planning activities; (2) the initiation, selection, and development of project ideas consistent with planning outcomes; (3) the evaluation of project results and the redirection of program plans; and (4) efforts to ensure that project results are used, with the consequent effect of improving military capability. (See Table 1.) These activities do not necessarily proceed sequentially, even for any one program area, but are often parallel efforts, highly interactive and oriented toward future conditions. In a sense, the program planning activities are efforts to anticipate where operational hardware will (or should) be in terms of advanced capability at some future period (e.g., 10 or 15 years hence), while the program development and evaluation activities are designed to see that the technology base moves forward at a sufficient pace to achieve this advanced capability. Finally, program use--or technology transfer for implementation in weapon systems--must be effected to make use of these improved (potential) military capabilities.

Program planning is accomplished in quite different ways in different federal R&D funding agencies (see Table 2). For example, the Goddard Space Flight Center develops very specific objectives in terms of building operational capabilities for the NASA missions; whereas in the National Institutes of Health, and in various research divisions of the National Science Foundation, program objectives concern the advancement of science in well-defined disciplines, often broken down into fields of inquiry. The latter are quite unlike the objectives that the Defense Advanced Research Projects Agency (ARPA) emphasizes for R&D programs for enhancing military capability.

---

[3]The subject matter addressed here draws heavily on the research undertaken by J. G. Wirt, A. Lieberman, and R. E. Levien, *R&D Management: Methods Used by Federal Agencies*, The Rand Corporation, R-1156-HEW, January 1974.

Table 1

A PROGRAM MANAGEMENT FRAMEWORK

1. Program Planning	a. Identifying future military and technological deficiencies. b. Assigning priorities to R&D objectives. c. Assessing problem scope, difficulty, cost, and avenues for solution. d. Preparing plans for analyses and program development.
2. Program Development	a. Generation of projects--analysis of program needs and solicitation of ideas for relevant resources. b. Selection of projects--evaluation of solicited and unsolicited proposals and allocation of R&D resources. c. Monitoring of projects--communication and direction to the research community relative to current program plans. d. Evaluation of projects--assessment, both current and in retrospect.
3. Program Evaluation (or Redirection)	a. Recurrent assessments of relevance, feasibilty, and cost/benefit throughout program life cycle. b. Plans for changes in response to realized and unmet progress. c. Tests of effects and cost following completion. d. Feedback of results to planning activities.
4. Program Utilization (or Technology Transfer)	a. Incorporating users' preferences in planning and design. b. Integrating the utilization and development processes through communication, collocation, and shared responsibility.

Table 2

EXAMPLES OF R&D MANAGEMENT PRACTICES

Agency	Program			
	Planning	Development	Evaluation	Utilization
National Aeronautics and Space Administration: Goddard Space Flight Center	Highly directed with specific objectives	Reliance on internal staff for project selection Close monitoring	Regularly and frequently evaluated	Plan for utilization included in projects
National Science Foundation: Research Applied to National Needs (RANN)	General objectives and priorities	Project ideas generated externally Consultation with research community on project selection Partial monitoring	Regularly evaluated, but infrequently	Plan for utilization included in projects
Engineering, Mathematics, Physics Divisions	By disciplines or fields of inquiry	Support of individual scientists External mail review with feedback to reviewers Little monitoring	Program manager defends his project selection decisions	Little direct attention
Biological and Social Sciences	By disciplines (within these, by lines of inquiry)	Support of individual scientists External panel review of borderline decisions Little monitoring	Evaluation implicit in considering renewed or continued support	Little direct attention
National Institutes of Health	By disciplines, collections of similar subjects	Support of individual scientists Unsolicited proposals Dual panel review (greatest attention to evaluating proposals) Little monitoring	Evaluation implicit in considering renewed or continued support	Little direct attention
Office of Naval Research	By disciplines and directed at attracting most capable researchers	Seek new researchers No formal review of project ideas Monitor activities (once or twice a year)	Evaluation implicit in considering renewed or continued support	A part of the program manager's mandate

SOURCE: Abstracted from research material developed through extensive interviewing by A. Lieberman and J. G. Wirt. See also J. G. Wirt, A. Lieberman, and R. E. Levien, *R&D Management: Methods Used by Federal Agencies*, The Rand Corporation, R-1156-HEW, January 1974.

There is also a broad range of techniques for developing R&D programs once plans have been articulated. Many of these rely on outside experts to evaluate proposals or to make judgments of one sort or another. For example, the NSF engineering, mathematics, and physics divisions rely on mail surveys; i.e., proposals submitted by scientists are sent out to independent experts with a request for a written appraisal of the proposals' scientific merit. Program directors then have the option of feeding back, to the same group of appraisers, the combined judgments to obtain further reactions. (Extensions of this concept are discussed in detail in Sections III and IV.) As for the evaluation and use of the various research programs, again practices vary; but often too little attention is paid during development to the ultimate application of results.

UTILITY OF TECHNOLOGICAL-ADVANCE ASSESSMENT

To further clarify areas where technological-advance assessment may make a difference, it is important for us to refine further what we mean by program *planning* and *development* activities. For example, some technological assessment techniques are better for planning for the reasonably short-term horizon or for evaluating projects for which proposals have already been submitted,[4] while others have been designed for the longer-term horizon.

Long-term program planning is particularly difficult. Technological-advance assessment may be very useful for this purpose--in gathering information (perhaps by making use of outside expertise) on a technology or program objective in order to identify important new prospects for R&D. These assessments may also indicate what resources will be required to pursue these new prospects. Such indications would be helpful in moderate-time-horizon planning, since they would provide guidance on the kinds of Requests for Proposals to issue and would focus very detailed evaluations of project ideas onto subjects having

---

[4]Evaluations at this stage of decisionmaking may best be conducted on the basis of scientific merit, novelty of approach, relevance to program objectives, and so forth.

potentially high payoffs. Early planning exercises may also make it easier to compare the applicability of R&D ideas under consideration for a variety of technologies, and the research community may be encouraged to propose projects for funding that the technology-advance assessment exercises reveal as most promising.

Since technological-advance assessments would obviously be useful to R&D planning and program development, how can such assessments be obtained? There are a number of issues that we must look at in some detail in deciding how to go about a technological-advance assessment. How do we identify the appropriate *scope* of our assessment problem? It is highly desirable to identify potential advances that will be militarily useful and also to consider different technologies that may be alternative solutions to the same problems. Once we have determined what our assessment objectives are, we ought to have some criteria for deciding *who* is qualified to make appraisals of potential advances in those objectives. Also, we must seriously consider the problem that some of the participants in an exercise may tend not to respond seriously and fully. In the early planning stages especially, discussion should not be at a level of detail that would pose a threat of revealing proprietary information to competitors. Finally, there are a number of questions of how best to use the various opinions that are obtained: how to condense them, analyze them, and input them to the decisionmaking process. These subjects are discussed in detail in Section III.

The concept of developing a technological base that will later be of use to advance the capability of operational military hardware rests on the premise that we can discern the potential military "worth" of future technological advances; this is ultimately the responsibility of senior DoD officials.

One approach to evaluating military worth is to develop scenarios for future time periods to be used in identifying the technological "weak links" in the systems our forces will need.[5] Such a study should

---

[5]A "top-down" planning of technological advancement was recommended by Dr. John S. Foster, Jr., in his keynote address for the national security session of the American Institute of Aeronautics and Astronautics convention, January 10, 1973.

identify a set of steps or decision points to consider in deciding whether advances in a given technological area would be worthwhile. This approach might be characterized as an effort directed at determining what a military service needs.

Another approach would be to survey the many different kinds of Required Operational Capabilities (ROCs) as identified by the individual services. To the extent that the underlying problems are similar, research and exploratory development may be warranted. The initiation of ARPA's human resources research could be viewed as arising out of such a perception. This approach would address the "use" issue, in that it would be likely to yield results the services *would implement*.

STRUCTURING A TECHNOLOGICAL-ADVANCE ASSESSMENT

A way in which technologies can be viewed as alternative solutions to the same future military objective is shown graphically in Fig. 1. After first looking at very broad planning objectives, we can identify the various technologies that contribute to these objectives. From these, we can determine what projects have been done in the past in each technological area and what future projects seem to be best fitted to advance the capabilities (or remedy the deficiencies) for the given objective.[6] The number of potential projects yielding some advancement of the overall planning objective could be quite large, both because of the number of technologies that contribute to any one objective and because of the number of options within a technological area.[7]

Statistical procedures and engineering analysis may both yield useful (and different) "figures of merit" for planning purposes. Analysis of aircraft turbine engines, for example, has yielded an equation interpretable in terms of a tradeoff surface among the desirable,

---

[6]This description is rather simplified, especially because it is highly desirable that the activities depicted in Fig. 1 *interact* with one another. Thus a first cut at the planning objectives may identify promising projects for funding; the results of these projects will provide valuable information for a second cut at the objectives, as well as an expanded technology base for possible utilization.

[7]Also, it is quite possible that a given project might benefit several objectives.

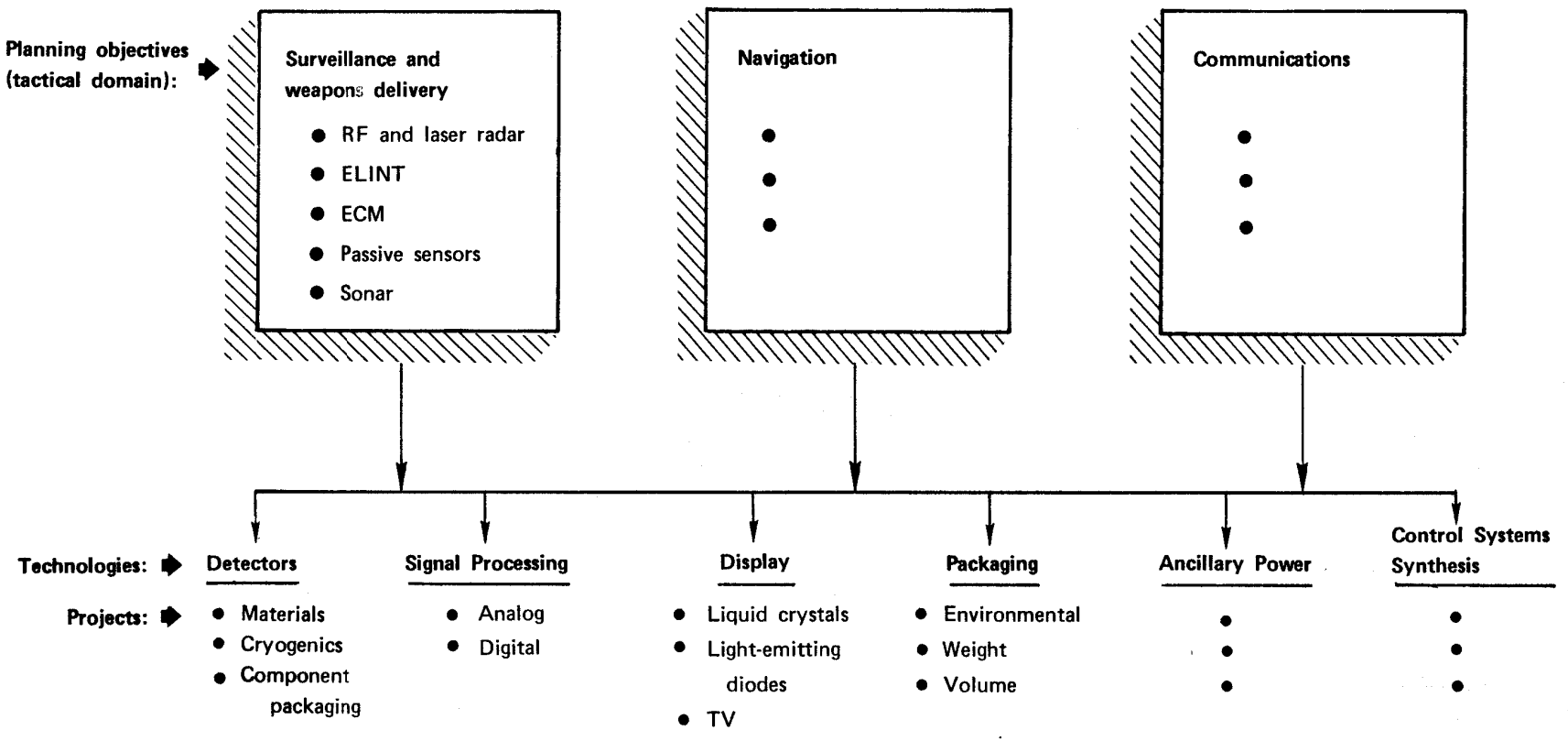


Fig. 1 — Structure for assessment problem

user-oriented performance dimensions. This tradeoff function captures, to some extent, both the feasible tradeoffs among performance dimensions from the standpoint of the underlying laws of physics and engineering-design considerations, and the degree of relative emphasis on the various performance dimensions that the *users* have sought. The overall equation is as follows:

$$\text{Engine "vintage"} = K \frac{\text{Temp}^{\alpha} \quad \text{Thrust}^{\beta} \quad Q \max^{\gamma}}{\text{Weight}^{\sigma} \quad \text{SFC}^{\eta}} ,$$

in which the engine "vintage" is a surrogate for the relative advancement of the various engines observed in past development activities, and the variables on the right are (in the numerator) the turbine inlet temperatures, the maximum thrust, and the pressure ratio; and the engine's weight and specific fuel consumption (in the denominator).[8] It should be emphasized that this kind of statistically derived relationship, which reflects in part the users' demands, is only appropriate for selecting new projects involving modest variations in each of these individual performance dimensions. For example, engines are not necessarily "superior" as we push for extreme reductions in specific fuel consumption, holding all other variables constant.[9]

Let us consider surveillance sensors for an illustration of the derivation, from engineering analysis, of a broadly based "figure of merit" that could be used to determine what technological areas we would want to include within a technological-advance assessment. Such

---

[8] For further details of this analysis, see Alexander and Nelson, op. cit. In the equation as actually estimated statistically, the engine "vintage" was captured by the time (in quarters since 1942) at which the engine passed its model qualification test. The various parameters were estimated as follows: $K = -1187$, $\alpha = 156$, $\beta = 18.8$, $\gamma = 11.7$, $\sigma = 26.5$, and $\eta = 20.6$.

[9] In fact, the estimated parameters would be misleading in any case for such combinations of variable values well beyond the range of the variables in the original data.

a figure of merit was constructed as a by-product of another recent research study.[10] In equation form, it is:

$$\varphi = \left(\frac{VW}{\Delta\delta^2}\right)R^2,$$

in which the figure of merit, φ, is the product of a measure of the information rate, $VW/\Delta\delta^2$, and the square of the standoff range, R. In this information-rate term, V stands for the attainable velocity at which the sensor moves and W stands for the swathe width; together the terms in the numerator are the measure of the area coverage rate. Δ is the intensity resolution of the system and δ is the spatial resolution.[11] One would also want to build into such a figure of merit the reliability of the system and its potential number of hours of operation in some standard calendar period (as a measure of the extent of all-weather and day-night capabilities). Such a figure of merit naturally leads to a broad investigation of appropriate technological advancement objectives; for example, V in part captures the data-processing rate of the surveillance sensor, and δ is determined by the capabilities of the front end of the sensor.

Such figures of merit could potentially be developed for many different military capabilities.

---

[10]See H. H. Bailey, *Side-Looking Radar* (U), R-581/3-ARPA, The Rand Corporation, June 1971 (Confidential).

[11]For further detailed description of these terms, see ibid., pp. 3-5.

III. COLLECTION OF GROUP JUDGMENT DATA: SURVEY DESIGN CONSIDERATIONS

In this section we will focus on the problems encountered in designing procedures for eliciting and grouping expert judgments. We believe that by eliciting the judgments of experts and studying their views on the relevant issues in a formal way (without necessarily looking for consensus), we will be able to improve on existing criteria for budget allocation that currently rely on a somewhat less formal methodology.[12]

The basis for our approach to assessing technological advance rests with the notion that there is such a thing as expertise, and that many experts are better than one. Here we will discuss some of the considerations surrounding this notion, and attempt to establish reasonable criteria for ultimately eliciting a collection of expert judgments on the same sets of questions. The basic considerations can be divided into four broad categories: the existence of expertise; the identification and selection of a panel of experts; the formalism and procedural questions associated with eliciting responses; and the design of a meaningful measuring instrument (a survey questionnaire). These categories are discussed below.

EXISTENCE OF EXPERTISE

In many instances it is difficult to argue that there is such a thing as expertise. For example, suppose it is 6 months before a national election and the question is, "Who will win the race?" No one really knows, and the knowledge of people who make a career out of studying elections is not significantly greater, at this time, than that of the average person. In another context, suppose we are

---

[12] One mechanism for eliciting the judgments of experts on fuzzy issues and then grouping them to obtain a consensus has been the Delphi Technique. The survey methods discussed below differ in many important respects from the conventional Delphi approach, as will become apparent.

interested in speculating about the "qualities of everyday life" in the year 2000 A.D. It is difficult to imagine that a few persons possess a greater degree of knowledge, intuitive understanding, and ability to predict such "qualities" than the rest of us. (In fact, if such greater knowledge does exist, it is not quite clear how it can be identified--but that is yet another kind of problem, and one that is considered below.)

The basic idea behind expertise is that there are people who have considerable knowledge and understanding of the mechanisms underlying particular problems and thus can do an appreciably better job of forecasting long-term trends and changes than the non-expert layman. The notion of visiting the oracle at Delphi to receive "expert" advice is an old one. We make the same mistake today of thinking that for every problem, there exists an expert problemsolver.

There is no doubt, however, that expertise does exist for some problems. A good physician, for example, can do a better job than a layman in assessing the likelihood that a given individual will develop cancer; similarly, it is easier for a good lawyer to predict the likely behavior of a judge or jury, in a given context, than a layman, and a good scientist or engineer can more accurately forecast technological change than someone without technical training and experience.

In the context of an R&D budget, once the planning objectives are carefully defined, the first step in an evaluation is to determine whether or not expertise really exists. If it does not, we must seek a new path for evaluation; if it does, we can proceed to the next set of considerations.

IDENTIFICATION AND SELECTION OF A PANEL OF EXPERTS

Having determined that expertise does exist for the problem at hand, we must now pick a panel of experts. A number of questions arise: How do we recognize an expert? How heavily should each opinion be weighted? How many experts constitute a "good" panel? Should all experts have the same type of expertise? How many panelists with each type of expertise should there be? Do the experts believe they are experts and how does that affect their judgment? What are the

common characteristics of experts? In selecting a panel, should their qualities be matched?

Attributes of a "Good" Panel of Experts

We believe that several characteristics are important for panels charged with assessing potential developments associated with technological change:

1. Diversity: Panel members as a group reflect a wide spectrum of talents. Good panels are not monolithic in terms of the fields of expertise represented. Rather, for every aspect of the problem under study, there is a panel member who is expert in that area. This representative diversity of disciplines is necessary so that the panel will not overlook, or give perfunctory treatment to, fundamentally important elements of the problem.
2. Depth: Some panel members have a profound understanding of the technical issues involved in certain problem areas. They are considerably more knowledgeable, in a scientific sense, than most people in the world, in their particular specialty. Thus, for every major scientific area that is a component of the basic problem, there is at least one panel expert who understands that component in depth.
3. Breadth: Good panels contain some members who are "systems experts," i.e., individuals who are accustomed to thinking on a broad level in terms of the interactions of various subsystems, of the implications of new subsystem developments on the economic feasibility of an entire system, and of the political, legal, social, and ecological aspects of the new development. Panel members who have a breadth of knowledge are better able to predict the feasibility and likelihood of a major technological breakthrough than, say, a "deeply knowledgeable" scientific expert who, in his narrower view, may ignore the need for developments in other fields that will make those in his own more meaningful.

There is no clearly defined "best mix" of experts for a given panel: What fraction of the panel members should be systems people and what fraction discipline experts cannot be precisely stated. But we can at least establish lower bounds. Once the problem has been broken down into some well-defined fields in which expertise exists, we know that there probably ought to be at least one expert from each field, and at least one systems analyst whose expertise has emerged out of that field.

Identification of Experts

A reasonable definition of expertise includes recognition and approbation by peer groups: a person is an expert in his field if others in his field consider him to be one. Some measures of expertise, by this definition, are the holding of office in a national scientific organization; a position on the editorial board of important technical journals in the field; awards for outstanding scholarship; honorary positions in national societies; publications of non-introductory books (monographs and advanced treatises); and awards of research contracts from various branches of the federal government. When a variety of professionals in a field are polled about whom they regard as an expert, and the same individuals are repeatedly mentioned, those individuals must be considered experts.

Outstanding systems analysts have typically been technical experts at one time and have later switched to administration for personal development. As a result, their perspective has broadened and their knowledge of related fields has increased, while their awareness of detailed developments in their original field has perhaps decreased. As they have interacted more with people from other disciplines, however, they have increasingly found instances where the dominating constraint on a development has involved some field other than their own. After coping with many diverse developmental efforts, these analysts have become adept at anticipating the feasibility, timing, and likely constraints associated with any new technological construct. They are usually employed as some type of manager (academic department chairman, research director, R&D manager for a corporation or a governmental agency, etc.).

Systems analysts are extremely valuable on a panel not only because of their broad perspective, but also because they tend to counterbalance the very conservative viewpoints typically found among scientific experts. These latter individuals, with their deep knowledge of a subject, have been indoctrinated throughout the years to exercise extreme caution (if not suspicion) about scientific breakthroughs and meaningful technological advances. They are not inclined to make intuitive judgments about broad issues, and so as a group they tend to be conservative about the feasibility, timing, and costs of new developments. Such a posture is "safe." If they are wrong, they won't be absurdly wrong and thus subject to criticism and loss of status. Systems analysts, on the other hand, tend to be less conservative. They will, for example, often substantially reduce the time estimates for a new development given to them by a scientific expert on their staff.

Selection of Panel Members

To minimize selection bias, panelists should be chosen by standard procedures developed in statistics[13] and the theory of psychological measurement.[14] It is easy to see how careless selection methods could reflect institutional rivalries and the personal biases of those who do the selecting. What is needed to start with is a listing of *all* known experts in each of the fields required for the analysis, and a similar listing for the systems people. This will establish populations of experts.[15] Then, after stratifying by field, simple random samples can be taken from each list (population). One alternative might be to stratify still further by preparing lists of experts in

---

[13] See, for example, M. W. Hansen, W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, New York: John Wiley & Sons, Inc., 1953.

[14] See, for example, American Psychological Association, *Standards for Educational and Psychological Tests and Manuals*, Washington, D.C., 1966.

[15] While it is conceivable that it may be desirable to stratify these populations by levels of expertise, one person's assessment is rarely appropriate.

government, industry, and the academic world for each field. We could then choose a simple random sample from each of these categories for each field. Systems analysts should also be chosen for the panel by stratified random sampling, using the same procedures. By using random number tables, in the usual way, to choose a random sample from these population lists, it should be possible to obtain judgments that are representative of those of the entire list, and to obtain the same results if the survey is carried out on several more occasions with similarly chosen samples. Moreover, the viewpoints or biases characteristic of a certain class of expertise (e.g., originating in industry) can be separately investigated and appropriate allowances made.

Motivation of Panel Members

A very important issue associated with panel selection concerns the motivation of the panel members to participate fully in the study. Suppose, for example, we have two experts in the same field who, for our purposes, are equivalent in expertise and we wish to compare their responses to a given question. If one expert gives an off-the-top-of-the-head response (a response based upon a few seconds or minutes of thinking and intuiting), while the other thinks through all the steps necessary to reach the final goal or development, evaluating the problems and constraints associated with each stage and assessing conditional probabilities, it seems reasonable that we should weight the careful judgment more heavily. If the panelists are not strongly motivated to cooperate fully, to the extent of providing careful, introspective responses, the results of the survey will not be reliable. Of course we can ask the panelist, on the questionnaire, how much time he devoted to preparing his responses; and we can word some of the technical questions so that the panelist is required to provide step-by-step responses as often as possible. But these approaches, while helpful, don't really come to grips with the source of the difficulty, namely, the panelist's lack of motivation.

Various means might be used to induce experts to respond cooperatively (assuming they agree to participate in the study in the first place). Possible motivational techniques include:

1. Honoraria: Participants could be paid a token fee. Such a fee could hardly be less than $50 or $100. If there were 30 panel participants, this would imply an honorarium cost of $1500 to $3000. But a token fee is too small to represent any real inducement to introspection; it is more likely to induce some panelists to agree to participate in the study with the intention of providing only $50 worth of effort. A significant fee of, say, $1000 or more could bring the cost of the study to $30,000 or more, which might be prohibitive. The larger fee, however, is much more likely to induce motivation than a simple honorarium. (Clearly, a $50,000 fee would induce almost any scientist to be very conscientious; but how much less will do the same job?)
2. Participant listing: The participants might be promised that they would be listed as a group in the written report of the study results. Such an approach is in keeping with the notion that seeing one's name in print is a strong inducement to making sure that the work associated with his name is substantial and correct. However, since the resulting report will not have the status of a paper in a prestigious scientific journal, nor will the individual be spotlighted as an innovator but will only be listed as one of a group whose opinion was sought, the motivating effect of such a listing is not likely to be significant.
3. Real-time interactions: The procedures by which information is elicited may influence its quality. A telephone interview by a technically competent interviewer can assess the seriousness with which the panelist is considering the questions and at the same time probe to determine if certain subtleties of the questions have been comprehended. However, the perspective of the interviewer himself may unduly influence the panelist. An alternative is real-time response and interaction among the panelists via a distributed computer system on which individuals would probe for the reasons behind each other's assessments as well as for characteristics of the group

of responses. This procedure might tend to encourage group self-motivation without the personal interactions often attributed to in-person panel discussions.

4. Research contract award potential: If the request to participate in the study comes from a potential source of research contract funds, the inducements to many individuals to cooperate in the study may be very great. This would be particularly true if all questionnaires were name-tagged and it was clear to every participant that the survey monitors were keeping track of how each individual was responding. Such a motivation may also lead to biases in the responses, as noted above.

PROCEDURAL ISSUES

Once an appropriate panel of respondents has been selected we must be concerned with some procedural and administrative issues. Should the questions of interest be asked in personal interviews with continual interaction and feedback between interviewer and interviewee? Should the questions be asked by telephone, or by mail? (There are advantages and disadvantages to each of these approaches.) Should some questions be addressed to panel members unknown to one another (so that their judgments cannot be impugned on an authoritarian basis)? Should some questions be addressed in group discussions, with all panel members freely interacting with one another, airing their views openly? Should there be one fixed set of questions or should the questions proceed in stages on one basis or another? Should panelists be asked the same questions repeatedly, after telling them the opinions of other respondents?

How can we ensure that the questions are valid, i.e., that they are phrased in such a way that they are really providing answers to the questions we want answered, with minimum semantical difficulties and maximum focus on the true points of interest? A well-known method for checking validity is to use the response results for forecasting and to compare these forecasts with actual outcomes. But when we are in a forecasting context in the first place, this approach may be

difficult to implement. A pilot study involving short-term prediction might prove helpful.

The choice of a correct administrative approach to be used in eliciting responses from a panel of scientific experts depends very much on the type of question being asked and on the use to which the responses will ultimately be put. It seems reasonable to expect that questions involving detailed scientific knowledge and expertise are best answered individually, with a minimum of outside bias from other panelists, from an interviewer, or from supervisory personnel. Other questions, such as those involving the potential use of policy variables that might greatly affect rates of technological change and development, are probably best handled in group discussions. For example, if a panelist knew that his particular research, while apparently not very important in and of itself, was in fact the major limiting factor to an extremely important development, and if he also knew that because of its importance, the funding level of support for his research might be increased 100 percent, his judgment about the feasibility and timing of some theoretical future development might be drastically altered. This type of background information, while difficult to supply in a questionnaire (since we can't always anticipate all the implicit questions asked, and the underlying assumptions made, by a respondent), is quickly requested and supplied in a group discussion.

The survey questionnaires might be administered by preceding their mailing with letters and or telephone calls advising that they are going to be mailed, followed by postmailing telephone calls[16] to check for their receipt and to clarify any questions. Such a procedure should help to ensure the validity of the questions and minimize non-response; it should also help to stimulate the respondents' introspection and cooperation.

Questions might be asked once, or they might be repeated in stages in a controlled way (see Section IV). There are many advantages to controlled feedback. For example, by requiring every panelist to

---

[16] In the basic mailing, panelists could be told that there would be a telephone followup.

provide some discussion (say, a paragraph of prose) about why he believes in his first-round position, and then later, why he is either adhering to this position or changing from it (after having been supplied with a complete set of first-round responses from all respondents), we are forcing all panelists to think through their judgments very carefully, vis-à-vis all other panelists, and we are zeroing in on the best rationale for the group judgments. Certainly no type of unanimous (or even consensus) judgment should be required for this type of analysis.[17]

Questions involving a paragraph of prose or an enumeration of reasons are useful for collecting ideas about how to regard an issue. We believe questions of this type are probably the most reasonable for the first stage of such a study. Later stages might involve more precisely focused questions to the same panel.

INSTRUMENT DESIGN

The design of a suitable instrument for measuring the judgments of scientific experts about the feasibility of some technological developments depends, of course, to a great extent on the specific developments we are interested in, the breadth and depth of the type of information sought, and the degree of detail required. However, at this preliminary stage, prior to actual design, we can at least establish some guidelines and general considerations that should prove useful.

Questionnaire design is an art about which much has been written.[18] There have been many guiding principles laid down on the basis of past experience that we cannot afford to ignore. They include the importance of giving the instrument a preliminary trial run in a pilot program, the care that must be exercised in wording questions, and the close

---

[17] Controlled feedback is also used in the Delphi approach (see Section IV). However, there it is typically used by asking panelists to justify their positions only if they are outside the interquartile range of the distribution of responses. Such an approach, of course, encourages agreement even when it may not be appropriate.

[18] See, for example, A. N. Oppenheim, *Questionnaire Design and Attitude Measurements*, New York: Basic Books, Inc., 1966.

attention that must be paid to the design of scales for recording judgments with quantitative content. Without commenting further on these important, but fairly standard, problems associated with all questionnaire designs, we will move on to discussion of some of the problems peculiar to our context.

Types of Questions

It seems appropriate, in our problem, that the survey questionnaire include at least three distinct categories of questions:

1. Questions dealing with the backgrounds of the individual panelists.
2. Questions dealing with the nature, format, and administration of the questionnaire.
3. Questions dealing with the scientific content motivating the entire study.

Questions in the first category are aimed at determining the panelist's level of expertise, his biases, the nature of his background (e.g., is he a systems manager?) and in general his qualifications for being on the panel. There might also be some questions directed at assessing his scientific conservatism. Asking for his judgment on a key scientific question, for which there already exists a known spectrum of viewpoints, might be a useful way of assessing his conservatism relative to that of the other panelists. The biases of a panelist might be revealed by asking for a listing of his best publications. Answers to this group of questions should provide some useful insights for understanding and interpreting the quality and perspective of each individual's responses.

Questions in the second category attempt to assess (a) the care that was exercised in completing the questionnaire, (b) whether or not the questions were clear, (c) how the question format might be improved in the next round, and (d) whether or not the panelist feels he was too constrained in his responses by the way in which the questionnaire was administered.

Questionnaire Goals

The goal of the survey questionnaire is to assist the manager in allocating his R&D budget. It is therefore desirable to condition the questions, whenever possible, on policy issues. For example, in asking a panelist to assess the feasibility of some new development, he might be asked for three separate assessments; one assuming a "low" funding level for the necessary research, a second assuming a "medium" funding level, and a third assuming a "high" funding level (of course, these terms need to be suitably defined either as part of the questionnaire, or as a part of the required response). It might be worthwhile to ask if a given development is likely within, say, 2 years, 5 years, 10 years, or perhaps never. It would also be useful to ask what other areas of research need to be "stimulated" because of their likely complementary payoffs. These other areas might only relate to peripheral aspects of the panelist's own work, or broaden its applicability.

Questionnaire Format

An important problem in this type of study is how to phrase the questions so that they cause the panelist to reflect thoughtfully about the problems and their constraints and limitations.

One set of questions might permit the panelist infinite latitude by asking him to enumerate all the steps that would be required to attain a given technological development. Another set, which might constrain him somewhat more, would provide him with some of the basic steps required to attain a given development, but would ask him to add or delete steps, as appropriate, and to assess conditional probabilities of being able to proceed down the chain of steps at each stage. The end result would be an "achievement tree" with many nodal points and assessments of the conditional probabilities of moving between any two nodal points. The degree of detail could be refined in successive stages of the questioning, as could the subjective probability assessments. Moreover, successive stages of questioning could easily lead to the emergence of completely new paths of development.

IV. QUANTIFICATION AND ANALYSIS OF GROUP JUDGMENT DATA

Research on quantitative assessment of technological advancement by the use of expert judgments appears to have had its formal genesis in studies carried out by Marshall and Meckling,[19] Klein,[20] and Summers,[21] who each made use of a quantity "A," the degree of technological advancement sought in a program. To estimate A, a sample survey was taken using four "experienced Rand engineers" as sample elements. The four subjects were asked to rate subjectively the magnitude of the improvement in the state of the art required for each of 22 aircraft and missile development programs. Their ratings were to be placed on a numerical scale ranging from 1 to 4. Group judgments were assessed by summing the ratings of the four subjects. After all the ratings were obtained, each program was categorized as small, medium, or large in the Marshall and Meckling and Klein studies. In the Summers study, the A variable (henceforth called the A-factor) and others were related to cost factors (ratios of actual to estimated costs) of a program by means of standard regression techniques.

Subjective assessments of the A-factor were attempted in two subsequent surveys. One was reported on by Harman and Henrichsen in 1970.[22] The sample subjects were asked to assess the A-factor for aircraft and missile systems on a scale of 0 to 20. As in the earlier

---

[19] A. W. Marshall and W. H. Meckling, *Predictability of the Costs, Time, and Success of Development*, The Rand Corporation, P-1821, December 1959, p. 17; also published in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, Princeton University Press, 1962.

[20] B. H. Klein, "The Decision Making Problem in Development," in R. R. Nelson, op. cit.

[21] R. Summers, *Cost Estimates as Predictors of Actual Weapon Costs: A Study of Major Hardware Articles*, The Rand Corporation, RM-3061-PR (Abridged), March 1965.

[22] A. J. Harman, assisted by S. Henrichsen, *A Methodology for Cost Factor Comparison and Prediction*, The Rand Corporation, RM-6269-ARPA, August 1970.

survey, the subjects were experienced Rand engineers, and the sample size for any given system ranged from two to four persons.

The last of the three surveys was made in 1970 in connection with the aircraft turbine engine.[23] In this survey, A-factors were assessed (among other things) for aircraft turbine engines on a scale of 1 to 20. The eleven subjects were all employees of the General Electric Company (some of the systems evaluated were manufactured by G.E. and some were not). Response was small and not all subjects made assessments for all systems.

Although the three surveys undertaken to assess A-factors have contributed considerably to our understanding of the problem of how to measure technological advancement, they still leave much to be desired. For example, samples have been so small that it is difficult to make meaningful statistical statements about the results. Also there are problems associated with asking individuals (no matter how expert) to compare objects having many characteristics or attributes on a single numerical scale. Not only will individual perceptions of reality tend to be different, but so will individual weights assigned to each attribute of an object being studied. Finally, individuals differ in their ability to quantify their judgments; even though they may all view an object in the same way, they may very well differ in their quantitative description of it.

This section discusses three separate methods of solving the problem of assessing technological advancement by quantifying sets of judgments: (1) a type of multidimensional scaling called "individual-differences scaling," (2) subjective probability assessment techniques, and (3) multivariate regression with categorical dependent variables. These methods have different properties, and each is useful for bringing out different types of relationships. Which method is most appropriate depends very much on the characteristics of the specific problem, the objectives of the study, and the form of the information available. For some problems, one particular method will be indicated, whereas

---

[23] A. J. Alexander and J. R. Nelson, *Measuring Technological Change: Aircraft Turbine Engines*, The Rand Corporation, R-1017-ARPA/PR, June 1972.

for another, each of the methods may be appropriate at a different stage of the study.

The method employing multivariate regression with categorical dependent variables is new and was developed specifically for this application. The mathematical description of the regression model in this section is only intended to be illustrative; a more extensive and general presentation appears in the appendix.

The individual-differences-scaling method is useful for placing the various systems to be compared in a multidimensional coordinate frame of reference so that the positions may be compared on each axis. Moreover, the characteristics of the assessors may be related to the underlying dimensions of the system attributes so that certain groups of assessors may be found to rate Project A more advanced than Project B on the first two dimensions, but not on other dimensions.

Subjective probability methods (the Bayesian approach) are useful for providing a formal mechanism for combining objective data with prior information (before observing the objective data); of course, in some cases there are no objective data available and we must rely solely on prior information. For example, cost relationships from other types of projects may be used for present projects even though the projects differ from one another. We may have cost information for one type of project, but not for another. Prior information may be introduced in the form of A-factors or in terms of probabilities. These approaches are compared later in this section.

In the multivariate regression method, a relationship is established between the probabilities that certain events will occur (or that certain propositions are true) and other attribute or explanatory variables related to these probabilities. For example, it may be of interest to relate the probability that a certain policy for funding R&D projects will result in a defined product, within a preassigned time, to explanatory variables such as funding rate, length of time for which the project is funded, type of project supervision (degree of closeness), and other variables. This sort of relationship can be established on the basis of a sample of previous data. Then, predictions may be suggested by the model.

The three methods, applicable to different types of problems, are discussed below.

MULTIDIMENSIONAL SCALING OF INDIVIDUAL DIFFERENCES

A recently developed and very powerful method of integrating collections of comparative judgments of individuals to form a composite group judgment, scaled on each of several dimensions, is called "individual-differences scaling."[24] The basic idea, applied in one context, is that each subject compares N projects regarding their relative degrees of technological advancement.[25] Thus, each individual renders $N(N - 1)/2$ judgments of the form: project S_i is more technologically advanced than project S_j, for all $i, j = 1, 2, \ldots, N$. In another context, each individual may order proposed R&D projects according to their probability of feasible development to a given stage by a given date. Next it is assumed that p dimensions are sufficient to represent the structure underlying the project differences. The ordinal judgments of a given individual can now be represented as ranks, or they can be converted into "distances" by one of several standard procedures, such as by the "law of comparative judgments."[26] Distances can be represented as weighted distances in Euclidian space, and the weights on each axis and the coordinates of each point can be estimated by the data. Thus, if the distance between project i and project j, as perceived by subject k, is

---

[24] J. D. Carroll and J. J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, Vol. 35, No. 3, September 1970, pp. 283-319.

[25] The definition of technological advancement to be used simultaneously for past, current, and future projects has not yet been made precise.

[26] See, for example, W. S. Torgerson, *Theory and Methods of Scaling*, New York: John Wiley & Sons, Inc., 1958; and R. D. Bock and L. V. Jones, *The Measurement and Prediction of Judgment and Choice*, San Francisco: Holden Day, 1968.

$$d_{ij}^{(k)} = \left[\sum_{t=1}^{p} w_{kt}(\chi_{it} - \chi_{jt})^2 \right]^{1/2},$$

i, j = 1,...,N, and k = 1,...,n, if there are n subjects who render complete sets of judgments. The w_{kt}'s and the χ_{jt}'s are estimated from the data. The result is a configuration of points in p-space representing the relative positions of the N systems as viewed by the n subjects, in total. The coordinates are the scale values (on a ratio scale) of each of the projects in each dimension. Thus, a collection of pairwise ordinal rankings for each of n individuals would not only yield a set of combined judgments of ratio-scale numerical values in p-dimensions for each project, but also a set of weights for each person in the sample, representing the importance that person places on each axis for each project.

The response of each subject may be thought of as being composed of an overall common mean response, plus a response effect reflecting his particular degree of expertise, plus a white-noise error term accounting for individual variations in ability to express one's state of knowledge (thus, two individuals with precisely the same state of knowledge would still differ in their responses because of their error terms). One implication of this type of reasoning is that since experts should produce responses that are closer to being correct than non-experts, great care should be exercised in selecting the subjects. Once a panel of potential subjects is screened for expertise, choosing a sample from this panel, which should be as large as possible consistent with cost and practicality considerations, will result in a small "individual-difference error."

The results of an individual-differences-scaling type of analysis might be used in several ways. One way would be to establish a functional relationship (by regression methods) between cost and the various dimensions of the projects. Then, the cost of a new project could be predicted by interpolation (extrapolation). Another use of the results might include changing the objectives sought in the new project

if it is found that coordinate values along a certain axis are unreasonable or intolerable. Finally, after studying the sets of estimated weighting factors obtained for each expert, a decision might be made to query certain individuals reqarding their weighting of a particular axis. This procedure might reveal important circumstances related to the R&D projects that were overlooked (or ignored) by other subjects. Perhaps subjects should be informed of the weights placed on each axis by others and then the entire process repeated.

SUBJECTIVE PROBABILITY METHODS

A-Factors versus Probabilities

Technological advancement might be measured in terms of A-factors, as described above, or it might be measured in terms of probabilities of some proposition. For example, if we speak of E_i as the proposition that project S_i will have, upon completion, A-factor A_i, $i = 1,\ldots,N$, then $p_i \equiv P\{E_i\}$ should be a monotonic function of the A-factor. That is, knowing p_i is equivalent to knowing the corresponding A-factor, and conversely. The real implication of this equivalence is that if the cost of a project, for example, is to reflect the degree of its technological sophistication, either A-factors or p_i's could be used as independent (exogenous) variables in the cost equations. It is not clear at this time which of the two is a better measure, in the sense that it can be better estimated and can therefore be used to generate better cost predictions. This point will be considered further below.

The relationship between the p_i's and the A-factors is monotonic but not unique: there are a large number of potentially useful and convenient monotonic functions. For example, a linear relationship is provided by

$$p_i = \frac{A_i - a}{b - a}, \qquad b > a,$$

where p_i denotes the probability that S_i will have A-factor A_i, and A_i is scaled on the interval [a,b], where a denotes the minimum degree of advancement and b denotes the maximum. This relationship is depicted in Fig. 2.

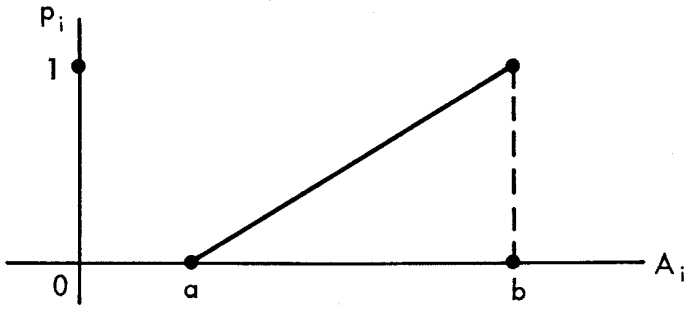


Fig. 2 — Linear correspondence between A-factors and probabilities

Another potentially useful functional relationship is the logistic correspondence,[27] given by

$$p_i = \alpha\left[\frac{1}{\beta + e^{-(A_i - a)}} - \frac{1}{\beta + 1}\right],$$

where $0 < \beta < 1$,

$$a^{-1} = \frac{1}{\beta + e^{-(b-a)}} - \frac{1}{\beta + 1},$$

$a \leq A \leq b$, and $a < b$. In this relationship, p_i is still a monotonically increasing function of A_i, but the function is convex or concave depending on whether $a < A_i < a - \log \beta$, or $a - \log \beta < A_i \leq b$, respectively. That is, there is a point of inflection at $A_i^* \equiv a - \log \beta$. The functional relationship is sketched in Fig. 3. The appropriate value of β might be selected with the use of regression techniques, after both the A-factor and the probability information are elicited.

Many factors enter into the problem of deciding whether probabilities or A-factors are better measures of technological advancement (in the estimation sense described above). Some of them are

---

[27] A multidimensional version of this form of correspondence is used explicitly, in a different context, in "Categorical-Dependent-Variable Multivariate Regression," below.

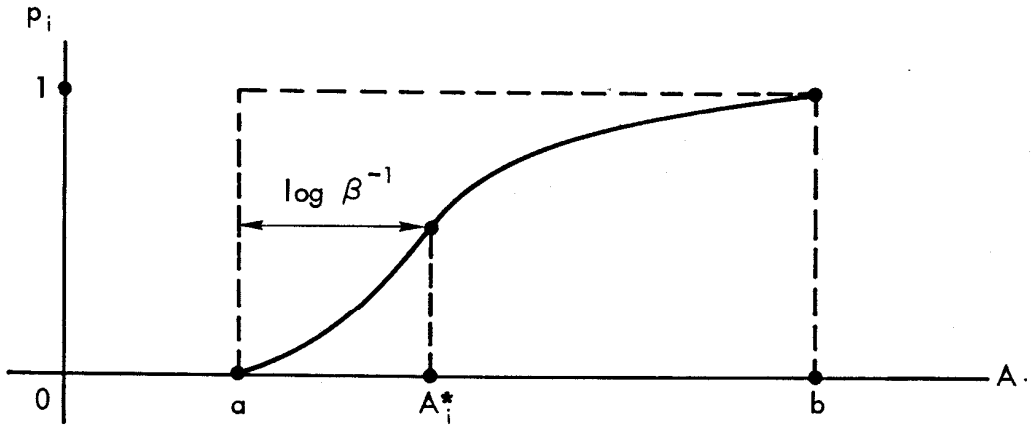


Fig. 3 — Logistic correspondence between A-factors and probabilities

1. One may be more *easily* assessable than the other because it may be easier to get subjects to render judgments of one rather than the other.
2. One may be more *precisely* assessable than the other in terms of repeatability of such assessment using different samples of "equally competent" experts; i.e., it may be subject to smaller variance.
3. One may be a more *accurate* measure than the other in terms of how close to the "truth" the measure can be expected to come (on the average) for a given sample size of independent and identically distributed observations.
4. One measure may be more *improvable* than the other, so that it may be possible, by feeding back information to experts over time, to teach them to be better assessors of one measure than the other.
5. Individuals may vary less in their *ability to quantify* their judgments about linearly scaled variables. There is also the difficulty that some subjects will have a better grasp of the meaning and concept underlying a probability than others. The ability of such people to make probability assessments of their judgments is keener than that of others, who might

possess the same substantive information regarding a given proposition but who are unable to quantify it as well.

Considerable research effort has been devoted to finding methods for getting subjects to assess scores for a variable, a difficult problem that has consumed much time among psychologists and statisticians.[28] The problem of assessing subjective probabilities has also received considerable attention not only in psychological and economic contexts,[29] but also in medical contexts.[30] However, whether probabilities or A-factors should be assessed remains an open question.

---

[28]See, for example, Torgerson, op. cit.; Bock and Jones, op. cit.; J. P. Guilford, *Psychometric Methods*, 2d ed., New York: McGraw-Hill, 1954; and C. H. Coombs, *A Theory of Data*, New York: John Wiley & Sons, Inc., 1964.

[29]A brief bibliography would include W. Edwards, "The Theory of Decision Making," *Psychology Bulletin*, Vol. 51, No. 4, 1954, pp. 380-417; W. Edwards, "The Reliability of Probability Preferences," *American Journal of Psychology*, Vol. 67, 1954, pp. 68-95; H. E. Kyburg, Jr., and H. E. Smokler, *Studies in Subjective Probability*, New York: John Wiley & Sons, Inc., 1964; M. G. Preston and P. Baratta, "An Experimental Study of the Auction-value of an Uncertain Outcome," *American Journal of Psychology*, Vol. 61, 1948, pp. 183-193; F. Mosteller and P. Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy*, Vol. 59, No. 5, 1951, pp. 371-404; R. L. Winkler, "Probabilistic Prediction: Some Experimental Results," *Journal of the American Statistical Association*, Vol. 66, 1971, pp. 675-685; R. L. Winkler, "The Quantification of Judgment: Some Methodological Suggestions," *Journal of the American Statistical Association*, Vol. 62, 1967, pp. 1109-1120; R. L. Winkler and A. H. Murphy, "'Good' Probability Assessors," *Journal of Applied Meteorology*, Vol. 7, 1968, pp. 751-758; H. J. Einhorn, "Expert Measurement and Mechanical Combination," *Organizational Behavior and Human Performance*, March 1972; L. J. Savage, "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association*, Vol. 66, 1971, pp. 783-801. For further references, see especially R. L. Winkler, "Probabilistic Prediction: Some Experimental Results," op. cit., and L. J. Savage, op. cit.

[30]See, for example, P. E. Meehl, *Clinical vs. Statistical Prediction*, Minneapolis: University of Minnesota Press, 1954, and L. W. Coppleson, R. M. Factor, S. B. Strum, P. W. Graff, and H. Rappaport, "Observer Disagreement in the Classification and Histology of Hodgkin's Disease," *Journal of the National Cancer Institute*, Vol. 45, No. 4, October 1970.

Bayesian Approach

The Bayesian approach to statistical inference and decisionmaking involves the assessment of prior distributions on the underlying parameters of the model.[31] One of the most important practical problems associated with the specific application of Bayesian analysis is that of how best to assess the prior distributions. But subjective probability assessors are very often "incoherent" in that their probability assessments for various related events are not consistent. For example, for some integer random variable E, an assessor might assert that as far as he is concerned, a priori, $P\{E > 0\} = 0.60$ and $P\{E > 10\} = 0.70$. One implication of these two assertions is that $P\{1 \leq E \leq 10\} = -0.10$, an absurd result.

It is expected that computers will be able to assist in the assessment problem. That is, routines could be developed to keep track of all previous assessments a subject has made about related propositions. The computer could ask the subject a sequence of questions designed to lead to consistent assessments of various propositions. If the responses are inconsistent, the computer would indicate this and request the subject to be more introspective and rethink his collective responses. Complete prior distributions could be assessed in this way. This type of computer-assisted assessment technique is a very realistic, potentially available development.[32]

---

[31] See, for example, D. V. Lindley, *Introduction to Probability and Statistics*, Vols. I and II, Cambridge: The University Press, 1965, for a discussion of Bayesian inference in univariate analysis, and S. J. Press, *Applied Multivariate Analysis*, New York: Holt, Rinehart and Winston, Inc., 1972, for its application in multivariate analysis. Some earlier work on the assessment of subjective probabilities, in the context of weapons development and the relation of subsystem properties to total system performance, may be found in F. S. Timson, *Measurement of Technical Performance in Weapon System Development Programs: Subjective Probability Approach*, The Rand Corporation, RM-5207-ARPA, December 1968.

[32] Some computer-assisted assessment techniques are already being used by M. Novick at the American College Testing Program, Iowa City, Iowa.

Controlled Feedback Methods

Technological advance might be assessed by some controlled feedback method (the Delphi method is one such approach) that seeks to obtain a consensus of a group of experts by feeding individual opinions back to the group so that they can revise their assessments. Each expert is typically subjected to a series of questionnaires. The summary statistics of the responses are fed back to the group, leading to a new round of revised responses. The feedback process is repeated several times. It is hoped that consensus will be achieved after several rounds. The experts are generally unknown to one another, and their opinions are often solicited by mail.

The Delphi method was originally devised in 1951 to apply expert opinion to the selection, from the viewpoint of a Soviet strategic planner, of an optimal U.S. industrial target system and to the estimation of the number of A-bombs required to reduce the munitions output by a prescribed amount.[33] Since then the original Delphi techniques and many variations of them have been applied (and misapplied) to a diverse collection of problems ranging from technological forecasting for advance planning for corporations, to studying national goals for the United States and for various foreign countries. Some of these applications are *not* ones for which expertise obviously exists (see Section III). After study and application of the techniques over the years, a collection of four summarizing reports appeared at Rand.[34]

---

[33]See N. Dalkey and O. Helmer, *The Use of Experts for the Estimation of Bombing Requirements: A Project Delphi Experiment*, The Rand Corporation, RM-727, November 1951, and N. Dalkey and O. Helmer, *An Experimental Application of the Delphi Method to the Use of Experts*, The Rand Corporation, RM-727 (Abridged), July 1962.

[34]N. C. Dalkey, *The Delphi Method: An Experimental Study of Group Opinion*, The Rand Corporation, RM-5888-PR, June 1969; B. Brown, S. W. Cochran, and N. C. Dalkey, *The Delphi Method II: Structure of Experiments*, The Rand Corporation, RM-5957-PR, June 1969; N. C. Dalkey, B. Brown, and S. W. Cochran, *The Delphi Method, III: Use of Self-Ratings To Improve Group Estimates*, The Rand Corporation, RM-6115-PR, November 1969; N. C. Dalkey, B. Brown, and S. W. Cochran, *The Delphi Method, IV: Effect of Percentile Feedback and Feed-in of Relevant Facts*, The Rand Corporation, RM-6118-PR, March 1970.

Other Rand studies have reported on more recent experimental results.[35] A critique of Delphi that focused on the misuse of controlled feedback procedures has been completed recently.[36]

The idea behind Delphi and other controlled feedback techniques is that if you want the best guess about a "fuzzy" question--one that is extremely difficult to answer even for the most informed people--ask an expert. Moreover, since for certain problems many expert heads are better than one, ask many experts and combine the conclusions, weighting them by the degree of expertise in the subject.

The details of how to implement this type of philosophy have varied from one application to another, but the basic idea remains the same. In some applications, the median response is fed back to the subjects for comparison with their own responses, and then a second-round response is sought. In other applications, upper and lower quartiles of the responses, as well as the median, are fed back. In some applications, subjects whose responses fall outside the upper and lower quartiles are requested to explain why they are outliers; otherwise, they are required to change their positions. In still other applications, subjects are required to extrapolate the future from earlier data, or they are required to provide paragraphs of prose describing their feelings or beliefs on an issue. It is the last-mentioned approach that we believe to be most appropriate, for reasons summarized in Section III.

Various theoretical and practical questions need to be raised and answered concerning the analysis of data collected in this way before we can seriously consider applying any controlled feedback technique to the problem of assessing technological advancement.

---

[35] N. C. Dalkey and D. L. Rourke, *Experimental Assessment of Delphi Procedures with Group Value Judgments*, The Rand Corporation, R-612-ARPA, February 1971, and N. C. Dalkey and B. Brown, *Comparison of Group Judgment Techniques with Short-Rand Predictions and Almanac Questions*, The Rand Corporation, R-678-ARPA, May 1971.

[36] H. Sackman, *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*, The Rand Corporation, R-1283-PR, April 1974. See also M. W. Hansen, W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, New York: John Wiley & Sons, 1953; American Psychological Association, *Standards for Educational and Psychological Tests and Manuals*, Washington, D.C., 1966.

For example, the distribution of responses in a Delphi study has been claimed to be approximately log normal.[37] However, since no cogent statistical evidence for this result has been found, it is believed that a more realistic interpretation of the observed data is that the distribution of Delphi responses is unimodal and skewed to the right, a behavior characterizing not only log normal distributions but many others, such as the gamma, as well.

Should the sets of responses to Delphi exercises be treated differently from one another, depending on the nature of the questions being raised? There is some evidence that they should.[38] Three basic types of questions can be distinguished: First, there is the "almanac" type of question, which has an easily established answer, but most people are not likely to know it. For verification, we might look for the answer in an almanac, or we might make direct measurements or observations (such as in the problem of guessing the weight of an apple). A second type of question involves group value judgments, such as determining the desirable quality of various aspects of life, or the desirable level of mean education that society should seek for its people. Dalkey and Rourke conclude that "Delphi procedures are appropriate for processing value material as well as factual material" (I believe this conclusion to be overdrawn). The third type of question requires the subject to make some predictions or forecasts. Dalkey and Brown carried out an experiment involving both almanac questions and short-range predictions (less than 6 months). They conclude that "no significant difference was observed between these two kinds of estimates." Again, there is reason to suspect such a conclusion, since the basis for such a significance test has not been indicated (was a log normal distribution of responses assumed, etc.?).

CATEGORICAL-DEPENDENT-VARIABLE MULTIVARIATE REGRESSION

In this section we will present a simplified version of a new methodological procedure, developed by Nerlove and Press, for doing

---

[37] See, for example, Dalkey, *The Delphi Method: An Experimental Study of Group Opinion.*

[38] See, for example, Dalkey and Rourke, op. cit., and Dalkey and Brown, op. cit.

regression analysis, using categorical dependent variables. We will show how the technique can be applied to the problem of assessing technological advance and of comparing the feasibility of competing R&D projects. The methodology generalizes the results of categorical-dependent-variable-regression, single-equation systems to correlated, multiple-equation systems of the same form. An extensive treatment of the subject is given in the appendix, and empirical applications of the method are presented elsewhere.[39]

Suppose there is a panel of n subjects, each member of which is asked to judge technological advancement for N projects, $S_1,\ldots,S_N$. Define the endogenous (dependent) indicator variables, $y_{ij}(k)$, as follows:

$$y_{ij}(k) = \begin{cases} 1, & \text{if event } E_{ij}(k) \text{ occurs,} \\ 0, & \text{otherwise,} \end{cases}$$

where $E_{ij}(k)$ denotes the event that in a pairwise comparison, the kth subject assesses the degree of technological advancement of S_i to be greater than that of S_j: $i, j = 1,\ldots,N;\ i \neq j;\ k = 1,\ldots,n$.

Next suppose the exogenous (independent) variables X_{1ij}, $X_{2ij},\ldots,X_{rij}$ bear directly upon whether project S_i is more technologically advanced than S_j. Let $x_{1ij}(k),x_{2ij}(k),\ldots,x_{rij}(k)$ denote the values perceived for $X_{1ij},\ldots,X_{rij}$ by the kth subject. These values really form the basis for the kth subject's assessments and may be thought of as quantitatively defining the set of assumptions he makes when he renders his comparative judgments.[40]

Adopt the model

$$y_{ij}(k) = F[\beta_{0ij} + \beta_{1ij}x_{1ij}(k) + \ldots + \beta_{rij}x_{rij}(k)] + u_{ij}(k),$$

---

[39] M. Nerlove and S. J. Press, *Univariate and Multivariate Log-Linear and Logistic Models*, The Rand Corporation, R-1306-EDA/NIH, December 1973.

[40] For example, the x's may measure the degree of information the kth subject has about the relative difficulty between two projects, or the depth of background or experience the kth subject has relative to projects i and j.

where $u_{ij}(k)$ denotes a random disturbance term with the properties that $E[u_{ij}(k)] = 0$, and $E[u_{ij}(k_1)u_{ij}(k_2)] = 0$, $k_1 \neq k_2$; and $F(\cdot)$ denotes a monotonic nondecreasing transformation with $0 \leq F(\cdot) \leq 1$; $i, j = 1,\ldots,N$; $k = 1,\ldots,n$. The function $F(\cdot)$ will be taken to be the cumulative distribution function of the logistic distribution, $F(t) = [1 + e^{-t}]^{-1}$, $-\infty < t < \infty$. The rationale for this choice is given in the appendix.

For algebraic simplicity, define the r-dimensional column vectors

$$\beta_{ij} \equiv (\beta_{0ij},\ldots,\beta_{rij})',$$

and

$$x_{ij}(k) \equiv [1, x_{1ij}(k),\ldots,x_{rij}(k)]',$$

so that the model becomes

$$y_{ij}(k) = F[x'_{ij}(k)\beta_{ij}] + u_{ij}(k),$$

for $i, j = 1,2,\ldots,N$; $k = 1,\ldots,n$. In any particular problem we assume that $x_{ij}(k)$ is given (subject k may specify it in part or in toto), and then $y_{ij}(k)$ is generated by assessment (when subject k renders his judgment about S_i vis-à-vis S_j).

Define

$$p_{ij}(k) \equiv P\{y_{ij}(k) = 1\} = P\{E_{ij}(k)\}.$$

That is, since

$$p_{ij}(k) = E[y_{ij}(k)],$$

$$p_{ij}(k) = F[x'_{ij}(k)\beta_{ij}],$$

$i, j = 1,\ldots,N$; $k = 1,\ldots,n$. Thus, by using the sample of n subjects to estimate β_{ij} as $\hat{\beta}_{ij}$, p^*_{ij}, the probability that S_i is more technologically advanced than S_j, given any preassigned set of assumptions, is estimated as

$$\hat{p}^*_{ij} \equiv F[x^{*\prime}_{ij}\hat{\beta}_{ij}], \qquad i, j = 1,\ldots,N.$$

But the disturbance terms $u_{ij}(k)$ are mutually correlated for various i and j, for a fixed k (even though they are assumed uncorrelated for different k's), so the $y_{ij}(k)$ are mutually correlated for fixed k. Hence, there is information in one equation that can be used, in part, to estimate parameters in other equations. That is, the system of equations should be viewed simultaneously as a set of multivariate nonlinear regression equations in which the endogenous variables (the y's) are reflective of a categorical probability and in which there is systematic heteroscedasticity.[41] The solution to this problem is a set of estimates of p^*_{ij}, for $i, j = 1,\ldots,N$, telling the analyst the relative degrees of technological advance required for a set of R&D projects (useful, for example, in cost equations) or telling the policymaker how to view the level of difficulty of a new project. The statistical method devised for solving this problem is given in detail in the appendix.

Suppose, for example, there are three systems to be compared regarding technological advancement. Recall that $y_{ij}(k)$ was defined to be one or zero, depending on whether or not in a pairwise comparison the kth subject assesses the degree of technological advancement of S_i to be greater than that of S_j, $i, j = 1,\ldots,N$, $i \neq j$, $k = 1,\ldots,n$. Take $N = 3$ (three projects to be compared) and define the random variable Y_{ij} for which $y_{ij}(k)$ is the kth observed value. The three random variables, Y_{12}, Y_{13}, Y_{23}, are mutually correlated and completely describe the comparative states of advancement of the three systems (note that if ties are ruled out, $Y_{ji} = 1 - Y_{ij}$ so that random variables

---

[41] Unequal variances of the disturbance terms for a given k.

Y_{ij} for which $j < i$ are unnecessary). Now rename the variables, $Y_1 \equiv Y_{12}$, $Y_2 \equiv Y_{13}$, $Y_3 \equiv Y_{23}$. Then, taking $Y_j = \theta_j + \varepsilon_j$, $j = 1, 2, 3$, where ε_j denotes an error term with mean zero, puts the problem into a three-equation model. The θ_j's will, of course, be taken to be the monotone transformations of linear combinations of independent variables discussed above. In this form the system can be thought of as a 2^3 or 2x2x2 (trivariate) contingency table that can, in turn, be thought of as a trivariate categorical-dependent-variable regression equation in which the dependent variable can assume eight possible values. If N systems are to be compared, there will be $N(N - 1)/2$ simultaneous equations to be solved in this way, rather than the three used in the example.

V. CONCLUSIONS AND RECOMMENDATIONS

We have seen how diverse procedures developed in different disciplines might be brought to bear on the problem of how to combine the opinions of individuals to form a group judgment about an ill-defined, multidimensional concept, such as the degree of technological advance required to complete a given R&D project, or the probability that a certain technological development will become feasible by a preassigned date. Although there are advantages with each procedure suggested, there are also various difficulties, uncertainties, and limitations, both conceptually and technically. The techniques described may also vary in their suitability for application according to the type of R&D management practices presently in use.

In summary, the methods of individual-differences scaling, subjective probability, and categorical-dependent-variable multivariate regression are very attractive for quantification and analysis of group judgment data. A controlled feedback approach could be useful in carrying out both an individual-differences scaling and a categorical-dependent-variable multivariate regression. That is, a panel of experts might evaluate the relative merits of a collection of R&D projects, perhaps both completed and projected. In the case of individual-differences scaling, group opinions can be fed back in a multistage approach to produce a scaled solution at each stage, because the current state of the methodology is strictly mathematical and has not yet advanced to the point where statistical inferences can be drawn.

In categorical-dependent-variable multivariate regression, there are advantages in being able to relate the experts' responses to their backgrounds, their degrees of expertise, and their definitions of the "difficulty" parameters of each project; also, statistical inferences and predictions are possible. However, at least at the present time, the method should not be used with feedback.[42] Nevertheless, the

[42] By using a multistage procedure that tells each subject at each stage what the group opinions were at the previous stage, we are in effect generating intrinsic collusion among the subjects. The effect

methodology may be used to advantage to analyze the first-stage results of a controlled feedback process.

Results derived from the assessment-data methods of analysis described above could be used to develop predictive models for determining the costs and feasibility probabilities of proposed and projected projects. Moreover, it should be possible to relate the cost and probability predictions to the assumptions underlying the predictions.

In conclusion, methods should be refined for selecting a panel of *appropriate* experts to provide assessments of technological advancement or feasibility of a collection of R&D projects. Planning objectives should be devised for selecting technology areas, and criteria should be established for determining the number and type of experts. If a set of projects is to be studied and evaluated, a controlled feedback type of analysis should first be pretested on the panel of experts. An interrogation procedure should be developed, with the questions designed to assess technological advancement in the specific technologies. Computer programs should be obtained or developed (if they are not accessible) for analyzing data by means of both individual-differences scaling and categorical-dependent-variable multivariate regression. At least two types of analysis should be undertaken: individual-differences scaling using a multistage controlled feedback approach, and a controlled feedback-based, categorical-dependent-variable multivariate regression but without the feedback data.

---

of such collusion upon the statistical analysis is to violate the assumption of independence among the subjects' responses. At the present time, the model formulation previously described will not permit correlated observations (interdependent response vectors). Hence, the results of a feedback process should not be analyzed in this way.

BIBLIOGRAPHY

Alexander, A. J., and J. R. Nelson, *Measuring Technological Change: Aircraft Turbine Engines*, The Rand Corporation, R-1017-ARPA/PR, June 1972.

Ament, R. H., "Comparison of Delphi Forecasting Studies in 1964 and 1969," *Futures*, Vol. 2, March 1970.

American Psychological Association, *Standards for Educational and Psychological Tests and Manuals*, Washington, D.C., 1966.

Bock, R. D., and L. V. Jones, *The Measurement and Prediction of Judgment and Choice*, San Francisco: Holden Day, 1968.

Brown, B., S. W. Cochran, and N. C. Dalkey, *The Delphi Method II: Structure of Experiments*, The Rand Corporation, RM-5957-PR, June 1969.

Carroll, J. D., and J. J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, Vol. 35, No. 3, September 1970.

Coombs, C. H., *A Theory of Data*, New York: John Wiley & Sons, Inc., 1964.

Coppleson, L. W., R. M. Factor, S. B. Strum, P. W. Graff, and H. Rappaport, "Observer Disagreement in the Classification and Histology of Hodgkin's Disease," *Journal of the National Cancer Institute*, Vol. 45, No. 4, October 1970.

Dalkey, N. C., *The Delphi Method: An Experimental Study of Group Opinion*, The Rand Corporation, RM-5888-PR, June 1969.

Dalkey, N. C., and B. Brown, *Comparison of Group Judgment Techniques with Short-Rand Predictions and Almanac Questions*, The Rand Corporation, R-678-ARPA, May 1971.

Dalkey, N., and O. Helmer, *The Use of Experts for the Estimation of Bombing Requirements*, The Rand Corporation, RM-727, November 1951.

-----, *An Experimental Application of the Delphi Method to the Use of Experts*, The Rand Corporation, RM-727 (Abridged), July 1962.

Dalkey, N. C., and D. L. Rourke, *Experimental Assessment of Delphi Procedures with Group Value Judgments*, The Rand Corporation, R-612-ARPA, February 1971.

Dalkey, N. C., B. Brown, and S. W. Cochran, *The Delphi Method, III: Use of Self-Ratings to Improve Group Estimates*, The Rand Corporation, RM-6115-PR, November 1969.

-----, *The Delphi Method, IV: Effect of Percentile Feedback and Feedin of Relevant Facts*, The Rand Corporation, RM-6118-PR, March 1970.

Edwards, W., "The Theory of Decision Making," *Psychology Bulletin*, Vol. 51, No. 4, 1954.

-----, "The Reliability of Probability Preferences," *American Journal of Psychology*, Vol. 67, 1954.

Einhorn, H. J., "Expert Measurement and Mechanical Combination," *Organizational Behavior and Human Performance*, March 1972.

Guilford, J. P., *Psychometric Methods*, 2d ed., New York: McGraw-Hill, 1954.

Hansen, M. W., W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, New York: John Wiley & Sons, Inc., 1953.

Harman, A. J., assisted by S. Henrichsen, *A Methodology for Cost Factor Comparison and Prediction*, The Rand Corporation, RM-6269-ARPA, August 1970.

Harman, A. J., *Choice Among Strategies for System Acquisition*, The Rand Corporation, P-4794, March 1972.

Klein, B. H., "The Decision Making Problem in Development," in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton: Princeton University Press, 1962.

Kyburg, H. E., Jr., and H. E. Smokler, *Studies in Subjective Probability*, New York: John Wiley & Sons, Inc., 1964.

Lieberman, A., and J. Wirt, *Management of R&D: Methods Used by Federal Agencies*, The Rand Corporation, R-1156, forthcoming.

Lindley, D. V., *Introduction to Probability and Statistics*, Vols. I and II, Cambridge: The University Press, 1965.

Marshall, A. W., and W. H. Meckling, *Predictability of the Costs, Time, and Success of Development*, The Rand Corporation, P-1821, December 1959; also published in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton: Princeton University Press, 1962.

Meehl, P. E., *Clinical vs. Statistical Prediction*, Minneapolis: University of Minnesota Press, 1954.

Mosteller, F., and P. Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy*, Vol. 59, No. 5, 1951.

Nelson, J. R., and F. S. Timson, *Relating Technology to Acquisition Costs: Aircraft Turbine Engines*, The Rand Corporation, R-1288-PR, March 1974.

Nerlove, M., and S. James Press, *Univariate and Multivariate Log-Linear and Logistic Models*, The Rand Corporation, R-1306-EDA/NIH, December 1973.

Oppenheim, A. N., *Questionnaire Design and Attitude Measurement*, New York: Basic Books, Inc., 1966.

Press, S. J., *Applied Multivariate Analysis*, New York: Holt, Rinehart and Winston, Inc., 1972.

Preston, M. G., and P. Baratta, "An Experimental Study of the Auction-value of an Uncertain Outcome," *American Journal of Psychology*, Vol. 61, 1948.

Sackman, H., *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*, The Rand Corporation, R-1283-PR, April 1974.

Savage, L. J., "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association*, Vol. 66, 1971.

Summers, R., *Cost Estimates as Predictors of Actual Weapon Costs: A Study of Major Hardware Articles*, The Rand Corporation, RM-3061-PR (Abridged), March 1965.

Timson, F. S., *Measurement of Technical Performance in Weapon System Development Programs: Subjective Probability Approach*, The Rand Corporation, RM-5207-ARPA, December 1968.

Torgerson, W. S., *Theory and Methods of Scaling*, New York: John Wiley & Sons, Inc., 1958.

Winkler, R. L., "The Quantification of Judgment: Some Methodological Suggestions," *Journal of the American Statistical Association*, Vol. 62, 1967.

-----, "Probabilistic Prediction: Some Experimental Results," *Journal of the American Statistical Association*, Vol. 66, 1971.

-----, and A. H. Murphy, "'Good' Probability Assessors," *Journal of Applied Meteorology*, Vol. 7, 1968.

Wirt, J. G., A. Lieberman, and R. E. Levien, *R&D Management: Methods Used by Federal Agencies*, The Rand Corporation, R-1156-HEW, January 1974.

Appendix

THEORY OF LOG-LINEAR AND LOGISTIC MODELS FOR THE ANALYSIS OF QUALITATIVE JUDGMENTS OF TECHNOLOGICAL ADVANCEMENT

Marc Nerlove and S. James Press

A.1. INTRODUCTION

In his excellent survey of the ideas that have been presented about inference in contingency tables, Mosteller (1968) stated, "I fear that the first act of most social scientists upon seeing a contingency table is to compute chi-square for it. Sometimes this process is enlightening, sometimes wasteful, but sometimes it does not go quite far enough." In this appendix we will summarize the most central earlier efforts, to give us an understanding of such cross tabulations, and then we will present some new results that promise to push our understanding a bit further.

This sequel is concerned with estimating relationships that determine jointly dependent qualitative variables, i.e., variables that are both categorical and unordered. For example, a military system requiring sophisticated, advanced technology may or may not be feasible by a preassigned date. We may wish to relate the probability that the new system will be feasible by that date to several explanatory variables, such as the rate of funding for the development of the required methodology, the probabilities that required subsystems will be available early enough, etc.

We will first provide a brief review of the single qualitative variable problem. Section A.2 treats the dichotomous, single qualitative variable case, and Section A.3 discusses the polytomous, single qualitative variable problem. Section A.4 generalizes the known results for a single qualitative variable ("univariate" case), treated earlier, to some new results for several jointly dependent qualitative variables ("multivariate" case). Results are generally applicable to a wide variety of problems involving the estimation of probabilities associated with R&D planning functions.

Equation numbers and figure numbers appearing in the appendix will always refer to those equations and figures appearing in the appendix.

A.2. ONE DICHOTOMOUS QUALITATIVE VARIABLE

FAILURE OF THE STANDARD REGRESSION MODEL AND ATTEMPTS AT FIRST AID

Let y denote a dichotomous random variable that is one if event E occurs, and zero otherwise, and let x: p x 1 denote a p-vector of variables used as explanatory or predictor variables for y (they may actually be transformations of more basic variables). Suppose (y_j, x_j), $j = 1,\ldots,n$ are independent observations of (y, x). Some of the explanatory variables may themselves be categorical and some may vary continuously. Let β: p x 1 denote a set of unknown weights so that $x'\beta$ specifies a linear function of x (the prime denotes transpose). Adopt the standard regression model

$$y_j = x'_j\beta + \varepsilon_j, \qquad j = 1,\ldots,n , \tag{1}$$

where ε denotes a disturbance term characterized by the distributional properties that $E(\varepsilon_j) = 0$; $\text{var}(\varepsilon_j) = \sigma^2$; $E(\varepsilon_j\varepsilon_k) = 0$, for $j \neq k$, $j, k = 1,\ldots,n$. If theory demands that the larger the value of $x'\beta$ the greater the chance that y will be one, we can think in terms of a monotone relationship between $x'\beta$ and the probability of event E (see Fig. A.1). The true probability function should generally have the characteristic sigmoid or S shape of a cumulative distribution function (cdf) since it must lie between zero and one and be nondecreasing. Thus, if $p \equiv P[E]$, p (which depends on $x'\beta$) is plotted as a function of $x'\beta$ as the solid curve in Fig. A.1. We have drawn the true curve so that its upper and lower asymptotes are indeed one and zero. Although the function can be well approximated in the center of its domain by the unconstrained dashed line segment AB, the unconstrained straight line is a poor approximation for very large or very small values of $x'\beta$, and, indeed, violates the condition that the function lie between zero and one for extreme values of the argument. A possible solution is to replace the dashed line by a broken line approximation that consists of the dashed line between the asymptotes and the asymptotes themselves above and below those values of $x'\beta$ for which the dashed line lies

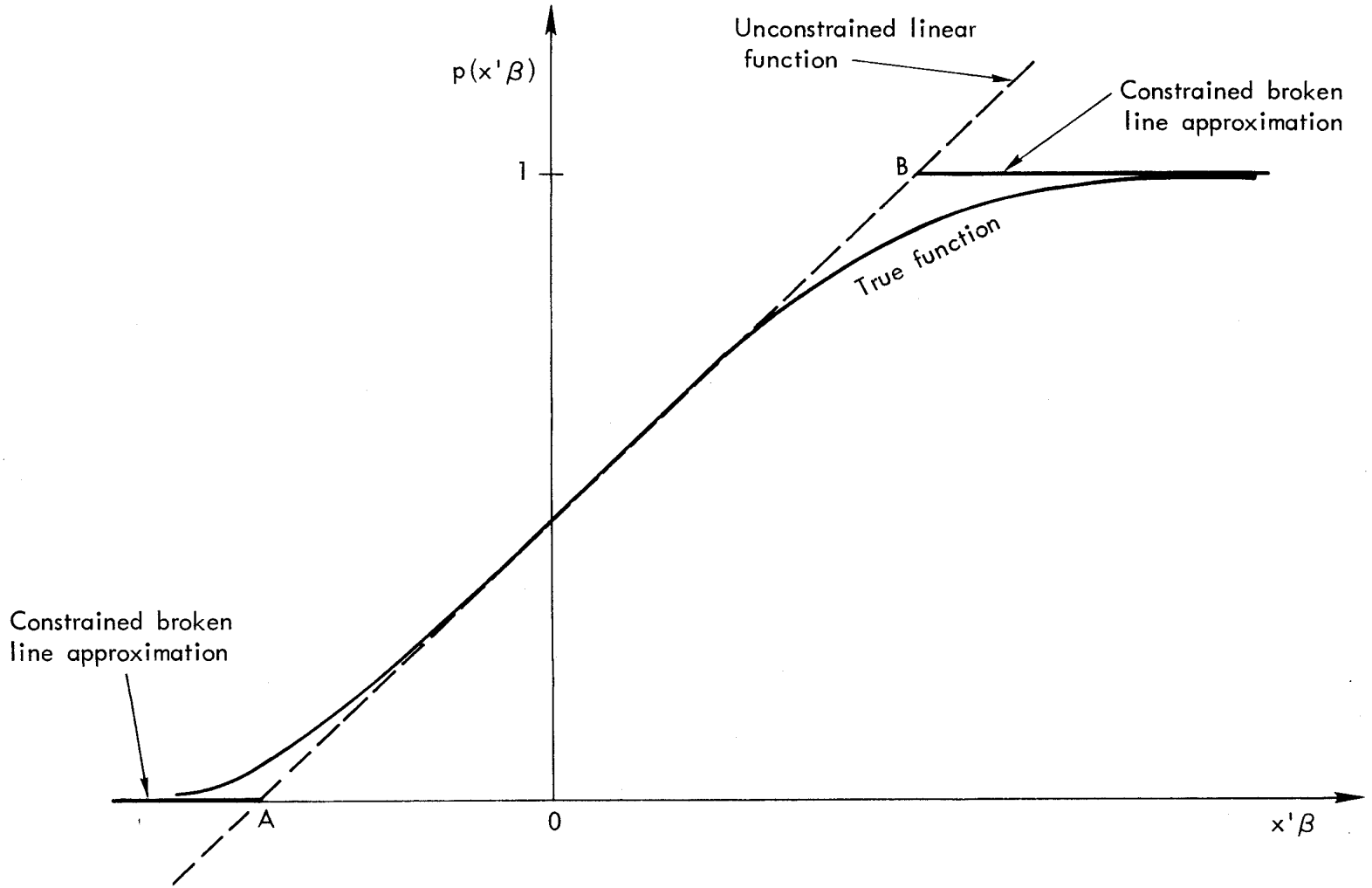


Fig. A.1 — Linear approximation to a probability function

outside the appropriate range. As can be seen from Fig. A.1, the broken line approximation is a good one except for two small regions near the points where the asymptotes are crossed. Of course, the size of such regions of poor approximation will vary depending on the shape of the true probability function.

Fitting a broken line approximation to binary observations resulting from a nonlinear probability function is no easy matter. If one had many observations of y for each value of x, as is typical in bioassay (see the next section), a number of solutions would be possible. Suppose, however, that there is only one observation of y for each value of x. This value may conveniently be called zero or one. Figure A.2 illustrates the effect of imposing the appropriate constraint on a broken line approximation. The observations are indicated by small circles lying on the upper or lower asymptotes of one and zero. It is possible to fit the data exactly, provided the observations do not force a negative slope on the interior segment. Indeed, the line of perfect fit is not even unique. If we were to impose the constraint that the interior segment be nonnegatively sloped, in the limit as the number of observations became very large, some values of $y = 1$ and of $y = 0$ would occur for values of x in inverted order, so the central portion of the optimal broken line segment would be vertical! In the multivariate case, difficult computational procedures would be required to fit a broken line approximation, but computational difficulties are the least of the undesirable features of such approximations.

Such difficulties have led many econometricians (see especially Goldberger, 1964, p. 249) to suggest the use of linear regression with the binary variable y as dependent. There are a number of important difficulties with estimating β in (1) by least squares. First note that under the assumptions of the model in (1), for fixed x_j, y_j is a Bernoulli random variable so that $E(y_j|x_j) = x_j'\beta$ and $\text{var}(y_j|x_j) = \text{var}(\varepsilon_j) = x_j'\beta(1 - x_j'\beta)$. Since $\text{var}(\varepsilon_j)$ depends upon j, the ε_j are heteroscedastic, and the use of ordinary least squares estimation will generate inefficient estimators and imprecise predictions. Note also that $x_j'\beta$ can have any value on the real line, whereas y_j, and therefore $E(y_j|x_j)$, is restricted to the unit interval. As a result, predictions

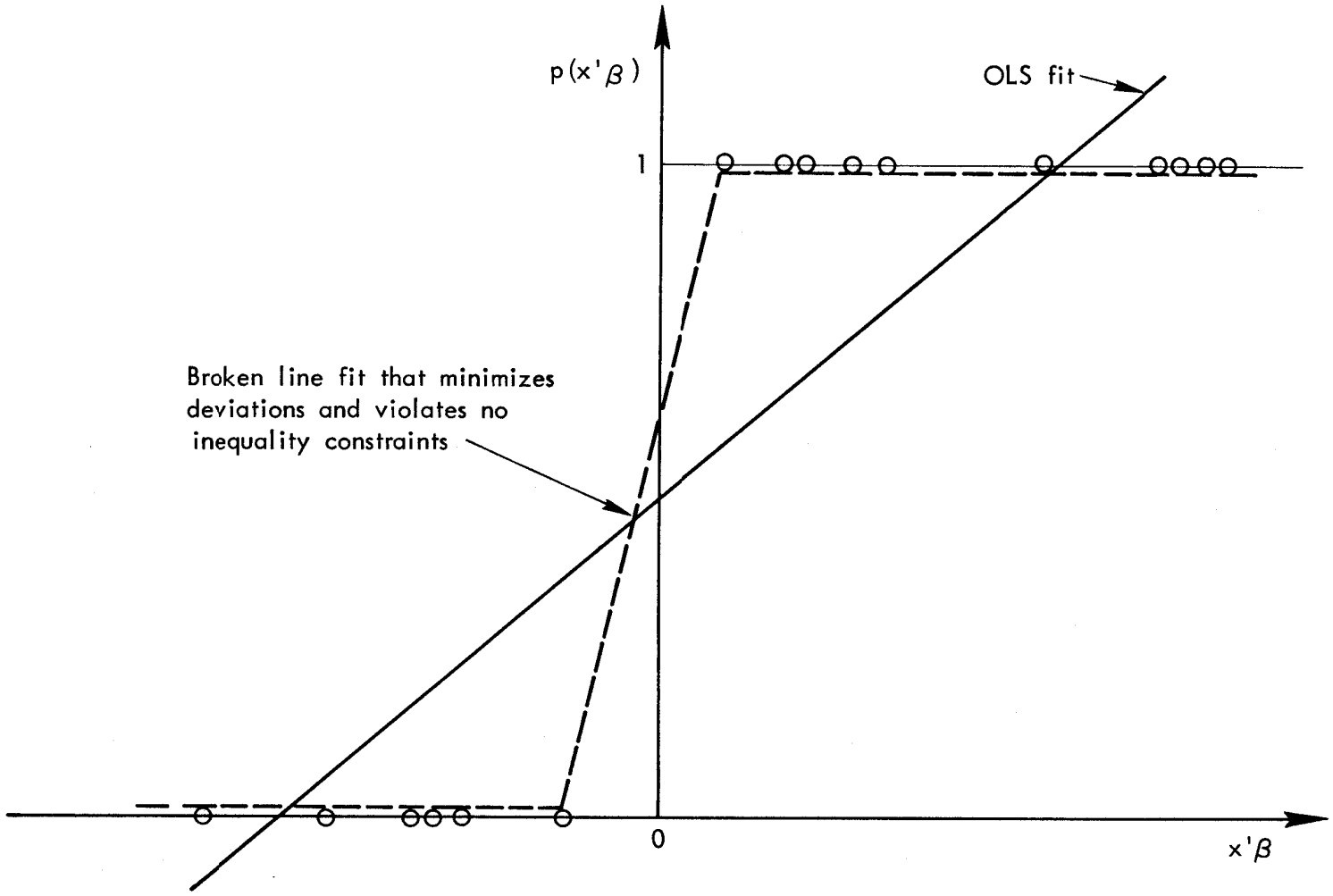


Fig. A.2 — Comparison of broken line and Ordinary Least Square (OLS)

lying outside the unit interval may sometimes be produced for small or large values of $x'\beta$. Note also that the fitted relationship is exceptionally sensitive to the location of the explanatory variables and that the usual tests of significance for the estimated coefficients do not apply. Further, multiple R^2 no longer is meaningful, and estimated standard errors are not consistent. Finally, we note, with Cox (1970, p. 17), that "because the y_j's are not normally distributed, *no* method of estimation that is linear in the y_j's will in general be fully efficient." That is, we can always improve on the least squares estimation (whether or not it is corrected for heteroscedasticity) since it is a linear estimator.

In an experimental situation, one would have control over many of the explanatory variables and could, and normally would, ensure that the values of these variables covered a wide range and were not abnormally bunched at high or low levels. In a nonexperimental situation, however, we are unlikely to be so fortunate as to have our explanatory variables distributed more or less uniformly over a wide range.

It has been suggested that the heteroscedasticity problem can be removed by using generalized least squares (see Goldberger, 1964, p. 248; Zellner and Lee, 1965, p. 387). Unfortunately, this approach can be relied upon only asymptotically. Moreover, except for making use of the first two moments of ε_j it completely ignores the Bernoulli character of the errors.

To correct for the problem of heteroscedasticity, Goldberger suggests replacing the variances of the ε_j in a generalized least squares analysis by the estimates $\hat{y}_j(1 - \hat{y}_j)$, where the $\hat{y}$ are the calculated values of the y from an ordinary least squares regression of y_j on x_j. Unfortunately, there is no guarantee that $\hat{y}_j$ will lie between zero and one for all j, so some of the "variances" may be negative.[1]

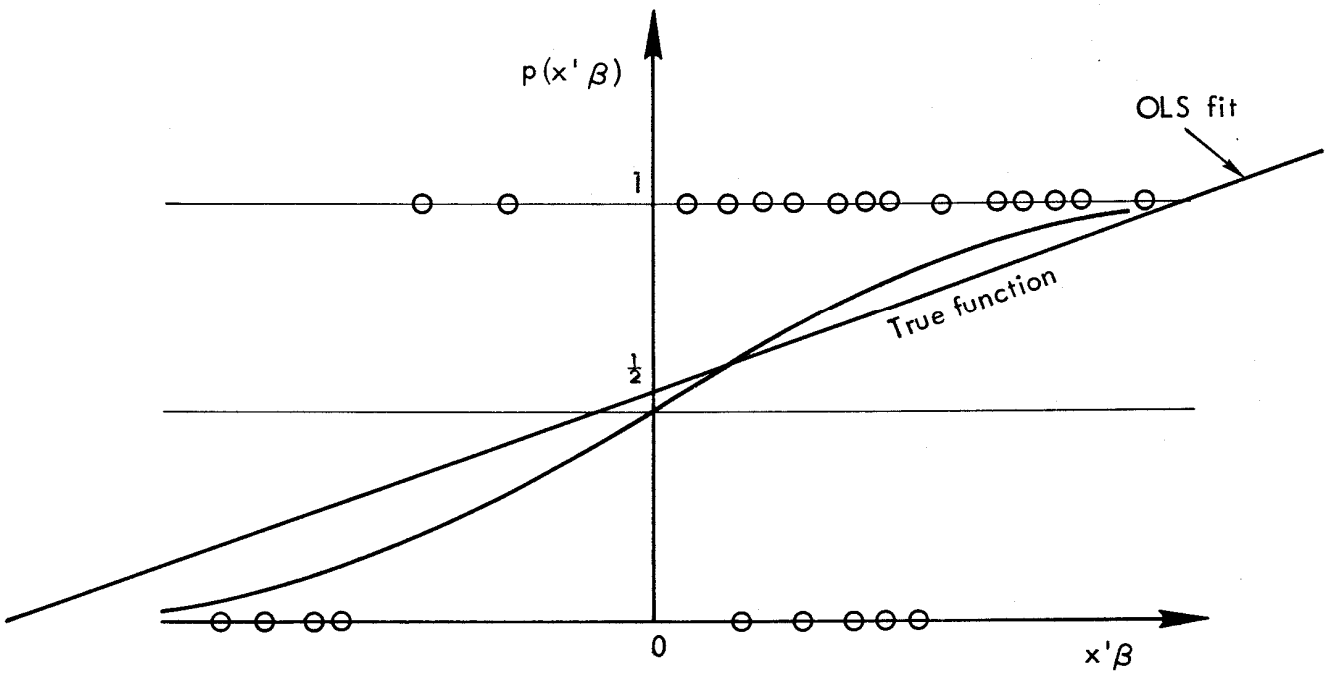
Smith and Cicchetti (1972) have done extensive Monte Carlo studies on alternative methods of handling inadmissible weights from a first

---

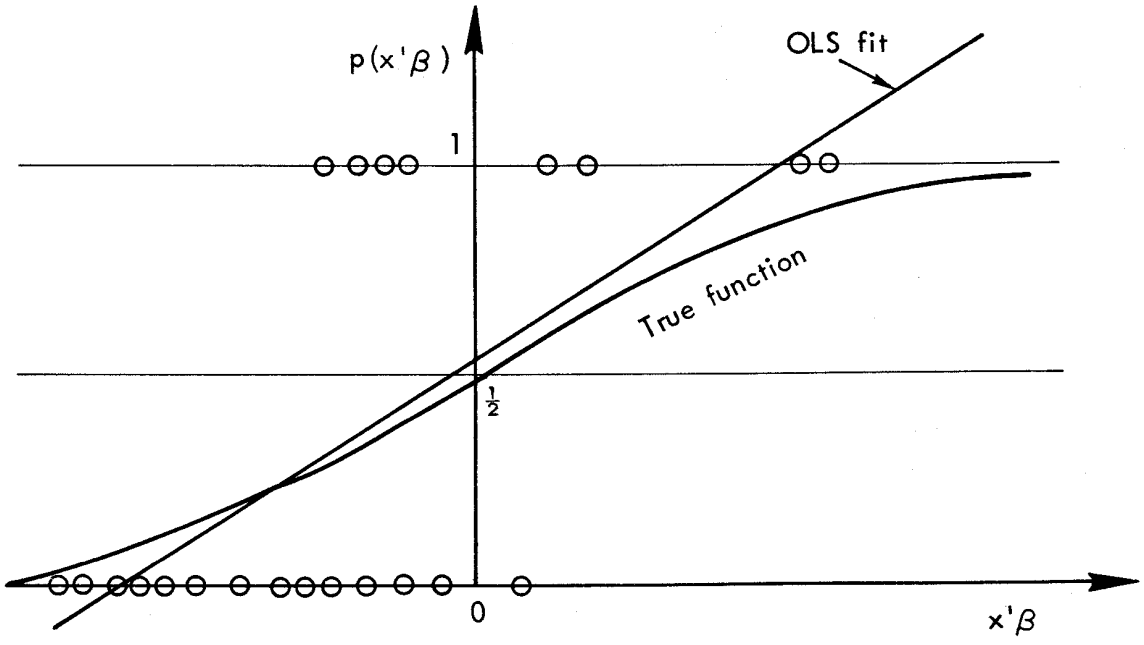
[1] McGillivray (1970) shows that $\hat{y}_j(1 - \hat{y}_j)$ is a consistent estimator of $(E\hat{y}_j)(1 - E\hat{y}_j)$, although the former may be negative but the latter cannot be.

stage OLS regression in a generalized least squares analysis. These are: (1) setting $\hat{y}_j = 0.5$ when $\hat{y}_j(1 - \hat{y}_j) < 0$; (2) replacing $\hat{y}_j(1 - \hat{y}_j)$ by $|\hat{y}_j(1 - \hat{y}_j)|$; and (3) replacing $\hat{y}_j$ by 0.98 when $\hat{y}_j(1 - \hat{y}_j) < 0$. These three generalized least squares estimators are compared with OLS for samples of size 20 and size 100 for regressions involving both dichotomous and continuous explanatory variables and several alternative sets of true coefficients. The results of these experiments are difficult to summarize except that the larger sample evidently reduces the apparent differences among all methods. All the estimates are biased, but the authors do not seem to have examined the relation of the dispersion of the explanatory variables to the amount of bias present. The estimated variances of the estimates are badly biased toward zero and "the power of the Student's t-test for testing hypotheses with the estimated coefficients is very limited."

Generating estimates and predictions that lie outside the unit interval can be avoided by doing "restricted" least squares estimation in which the estimators are constrained to the unit interval in the minimization. Especially when heteroscedasticity is accounted for, the resulting computation is particularly complex, it is reliable only in large samples, and it ignores most of the distributional properties of the errors. Figure A.3 illustrates the sensitivity of an unrestricted least squares fit of a linear probability function (Eq. (1)) to bunched data. When most values of $x'\beta$ are large, the preponderance of observations on y will be 1; when most of the values of $x'\beta$ are small, the preponderance will be 0. The location and slope of the fitted relationship may be greatly affected. In an extreme situation (when all values of $x'\beta$ are zero or all are one), the fitted relationship may have zero slope and no intersection with the ordinate between zero and one. No method of estimation can possibly function well in such extreme cases, but the linear probability function is more likely to be subject to such difficulties in non-ideal situations than a method more closely approximating the true probability function. Misspecification is bad; bunched data are generally bad; combining one with the other can only aggravate the difficulties.



(a) Many high values of x'β



(b) Many low values of x'β

Fig. A.3 — Illustration of the sensitivity of Ordinary Lease Squares (OLS) to the values of the explanatory variable

PROBIT ANALYSIS (MANY OBSERVATIONS PER CELL)

It was pointed out above that applications of standard regression techniques to binary data has many technical difficulties. One of the earliest methods of analyzing such data, "probit analysis," has been attributed to J. H. Gaddum and C. I. Bliss by Ashton (1972). Some of this early work was described in Bliss (1934). Finney (1947) applied probit analysis to the problem of analyzing quantal (binary) responses in bioassay.

The bioassay problem is one of studying, for example, the potency of some drug by observing the behavior of animals to whom various doses of the drug have been administered. Suppose each animal in a group of n_j animals is given a dose t_j of a poisonous drug, $j = 1,\ldots,G$; let y_{ij} be unity if animal i in group j dies, and let y_{ij} be zero otherwise, $i = 1,\ldots,n_j$. Let p_j denote the probability that an animal in group j will die from the drug, and let $(a + bt_j)$ denote the "stimulus level" of the drug, where a and b are unknown constants. If one wants a relationship in which p_j is a nondecreasing function of t_j, the model

$$p_j = F(a + bt_j) ,$$

where $F(t)$ denotes a cumulative distribution function, may be used; so that automatically $0 \leq p_j \leq 1$ and p_j is a nondecreasing function of t_j for $b > 0$. More generally, let $x_j'\beta$ denote the stimulus level, where x_j: $p \times 1$ denotes a p-vector of stimulus variables and β denotes a p-vector of weights. Then the model is

$$p_j = F(x_j'\beta), \qquad j = 1,\ldots,G , \tag{2}$$

where p_j is nondecreasing in $x_j'\beta$, but may be decreasing in some stimulus variables, depending upon the algebraic signs of the components of β (the elements of x may be transformations of some more basic variables).

Probit analysis describes a collection of methods for estimating the relationship in (2), using grouped data,[2] when $F(t)$ is taken to be

[2]The term "grouped data" is used throughout synonymously with the term "many observations per cell," to mean that there is more than one observed y (dichotomous or polytomous) for each observed x-vector.

the cdf of the standardized normal distribution. First define the sample estimator of p_j,

$$\tilde{p}_j \equiv \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}, \qquad j = 1,\ldots,G , \tag{3}$$

which is seen to be a reasonable estimator since $p_j = P\{y_{ij} = 1\}$; it is also a maximum likelihood estimator. Next define

$$z_j \equiv \text{Probit } (\tilde{p}_j) \equiv \tilde{t}_j + 5 , \tag{4}$$

where $\tilde{t}_j$ is defined by $\tilde{p}_j = F(\tilde{t}_j)$; that is,

$$\tilde{p}_j \equiv \int_{-\infty}^{\tilde{t}_j} \frac{1}{\sqrt{2\pi}} e^{-1/2t^2} \, dt .$$

The 5 is added to the fractile $\tilde{t}_j$ in (4) to avoid negative values for z_j. Finally, if we adopt the model

$$z_j = x_j'\beta + \varepsilon_j, \qquad j = 1,\ldots,G , \tag{5}$$

where $E(\varepsilon_j) = 0$, $\text{var}(\varepsilon_j) = \sigma^2$, $E(\varepsilon_j\varepsilon_k) = 0$, $j \neq k$, β in (5) may be estimated by ordinary least squares. For this probit analysis method to be useful, there should be several observations per cell ($n_j >$ for every j). Moreover, efficiency of estimation is lost in the *ad hoc* procedure associated with the added 5 in (4). Note also that there are computational difficulties associated with the use of the integrals in this procedure, unequal numbers of observations per cell are inefficient, and cells with one or zero observations per cell are not useful.

The use of probit analysis in econometrics and the social sciences has not been widespread. Economic and sociological data are generally nonexperimental (biological data are usually the result of designed experiments), and it has until recently been rare to collect more than one observation per cell.[3] In such cases, frequency estimates of the

[3]For an exception, however, see Tobin (1955).

type given in (3) have not been possible. Now, however, increasing availability of survey data for individuals, households, or firms makes methods that rely upon frequency estimates more acceptable and more usable than ever before.[4]

LOGIT ANALYSIS (MANY OBSERVATIONS PER CELL)

Logit analysis is a term coined by Berkson (1944, 1953, 1955) to characterize methods of estimating the relationships in (2), using grouped data, when F(t) is taken to be the cdf of the standardized logistic distribution; that is,

$$F(t) = \frac{1}{1 + e^{-t}} , \quad -\infty < t < \infty . \tag{6}$$

Combining (2) and (6) gives

$$p_j = \frac{1}{1 + e^{-x_j'\beta}} . \tag{7}$$

Solving for the argument gives

$$x_j'\beta = \log\left(\frac{p_j}{1 - p_j}\right) . \tag{8}$$

Berkson defined the estimated log-odds,

$$z_j \equiv \log\left(\frac{\tilde{p}_j}{1 - \tilde{p}_j}\right) \equiv \text{Logit } (\tilde{p}_j) , \tag{9}$$

where $\tilde{p}_j$ is defined in (3). Now adopt the model of (5) and estimate β by least squares (note that $\tilde{p}_j$ cannot be zero or one in (9), which would occur if $n_j = 1$). It is well known that the numerical difference between the normal and logistic cdf is very slight except at either of

[4] This point has been stressed much earlier by Orcutt, Greenberger, Korbel, and Rivlin (1961).

the extremes (see Cox, 1970, p. 28; or Ashton, 1972, p. 11). The logit analysis method has the advantage of numerical simplicity over probit analysis, but the estimation method using logits still requires the availability of grouped data.

The generalized least squares method suggested by Theil (1970, p. 109) for analyzing binary response data for a single qualitative variable deletes cells containing only one or no observations and is applicable only in large samples. Unfortunately, in an economic context it is easy to envision situations in which *every* cell will have only one or no observations (for example, when one of the explanatory variables is not under control and is continuous) and in which the sample is small. A continuous variable can always be categorized, but unless the total number of observations is very large, much information may be lost. It is still desirable to develop methods that are directly applicable to one observation per cell and will have credible small sample properties. The method of maximum likelihood, described below, has such properties and was applied to this problem at a very early stage.[5]

---

[5] An alternative method of estimation called "minimum Chi square" was devised by Berkson (1955). The results are asymptotically identical to maximum likelihood estimation (MLE) but in large samples are simpler to evaluate. The procedure is developed from the idea of minimizing the conventional Chi square statistic for a multinomial table,

$$\chi^2 = \sum_{j=1}^{G} n_j[\tilde{p}_j - F(x_j'\beta)]^2\{F(x_j'\beta)[1 - F(x_j'\beta)]\}^{-1}$$

with respect to β. The quantities $\tilde{p}_j$ and $F(t)$ are defined in (3) and (6). This result is equivalent to weighting the squared deviations between observed and expected numbers of observations by the reciprocals of the variances of the observations. When large samples are present, this expression may be linearized to yield very simple solutions for the elements of β. When all y_{ij} in a cell are zero or one, however, the results must be modified appropriately so as not to lose this information. In any case, the minimum Chi square procedure requires more than one observation per cell (preferably many observations per cell).

The logistic transformation has been justified for studying relationships of the form given in (2) on the basis that it is simple and tractable algebraically, and therefore simple computationally, and its cdf is numerically close to the normal cdf. These facts, though useful and interesting, can hardly be considered adequate justification for the arbitrary selection of a functional form for F(t). In a more fundamental justification, Truett, Cornfield, and Kannel (1967) provide a physical mechanism for this useful transformation. Although their work was carried forth in the setting of a disease and its associated syndrome, the ideas apply generally.

Let y = 1 if event E occurs, and y = 0 if E does not occur--that is, if $\bar{E}$ occurs. Then, if x is a p-vector with continuous density $h(x|\Theta)$, where Θ is a parameter matrix indexing the distribution, by Bayes theorem,

$$P\{E|x\} = \frac{P\{E\}h(x|E,\Theta)}{P\{E\}h(x|E,\Theta) + P\{\bar{E}\}h(x|\bar{E},\Theta)} .$$

Let $q \equiv P\{E\}$, and let $p \equiv P\{E|x\} = P\{y = 1|x\}$. Then,

$$p = \frac{1}{1 + \left(\frac{1 - q}{q}\right)\frac{h(x|\bar{E},\theta)}{h(x|E,\theta)}} .$$

Now suppose that given (E, Θ), $\mathcal{L}(x) = N(\theta_1, \Sigma)$; and given $(\bar{E}, \Theta)$, $\mathcal{L}(x) = N(\theta_2, \Sigma)$.[6] Then, it is straightforward to check that p may be written in the form

$$p = \frac{1}{1 + e^{-x'\beta}} ,$$

for some vector of weights β that depend upon θ_1, θ_2, and Σ; i.e., p follows the logistic cdf form.

---

[6] $\mathcal{L}(x) = N(\theta, \Sigma)$ implies that the probability law of x is normal with mean θ and covariance matrix Σ.

It may be argued that in many situations x does not follow a multivariate normal distribution. Halperin, Blackwelder, and Verter (1971) compared conditional MLEs for this model, assuming some of the x were dichotomous, with discriminant function estimates obtained assuming normality of the x (the latter are unconditional MLEs). Their results showed that when normality of the x was violated, ML estimators still behaved appropriately (and gave somewhat better fits to the model than discriminant function estimators). For this reason, as well as for those mentioned earlier, we restrict attention to logistic transformations of the data.

OTHER TRANSFORMATIONS WITH MANY OBSERVATIONS PER CELL

Transformations other than the normal and the logistic have been proposed as appropriate functions in (2) for studying relationships like those discussed in this report. Several of them, for example,

$$F_1(t) = 1/2\{1 + \sin t\}, \qquad -\pi/2 \leq t \leq \pi/2 ,$$

$$F_2(t) = 1/2 + \frac{1}{\pi} \arctan t, \qquad -\infty < t < \infty ,$$

$$F_3(t) = \tanh t/2, \qquad -\infty < t < \infty ,$$

are often grouped together under the designation "angular transformations." They are numerically similar to the normal curve, except in the tails. Moreover, in some econometric problems where there is primary interest in one or both of the tails of the distribution (rare or very frequent events) an angular transformation may prove to be more appropriate than a logit or probit. In such a case the parameters of the model may be estimated by least squares, provided that there are many observations per cell. (If not, then maximum likelihood methods similar to those developed in this report must be formulated.)

For example, using $F_1(x_j'\beta)$, if $z_j \equiv \arcsin(2\tilde{p}_j - 1)$, where $\tilde{p}_j$ is given in (3), and ε_j denotes a disturbance term, one may adopt the model

$$z_j = x_j'\beta + \varepsilon_j ,$$

and estimate β by least squares. $F_2(t)$, $F_3(t)$, and other transformations may be adapted in an analogous way.

MAXIMUM LIKELIHOOD ESTIMATION

Estimation of the model in (7) by maximum likelihood is well established for the logistic cdf case (see Cox, 1970, pp. 87-91; Hodges, 1958; Berkson, 1955), and also for the normal cdf case (see Tobin, 1955 and 1958 for an application in economics; Fisher, 1935; Finney, 1947; and Cornfield and Mantel, 1950). The general idea is outlined below.

The likelihood function for the model in (2) is

$$L(y_{11},\ldots,y_{n_G G}|x_1,\ldots,x_G) = \prod_{j=1}^{G}\prod_{i=1}^{n_j}[F(x_j'\beta)]^{y_{ij}}[1 - F(x_j'\beta)]^{1-y_{ij}}. \quad (10)$$

For problems in the social sciences where there is often just one observation per cell, $n_j = 1$ for all j, we can take $y_j \equiv y_{1j}$, and rewrite the likelihood function as

$$L(y_1,\ldots,y_G|\cdot_i,\ldots,x_G) = \prod_{j=1}^{G}[F(x_j'\beta)]^{y_j}[1 - F(x_j'\beta)]^{1-y_j} . \quad (11)$$

In either case, (10) or (11) may be maximized, usually by numerical methods, once a suitable $F(\cdot)$ transformation has been selected. In the remainder of this report (see above, p. 56 ff, for a justification) we focus upon the logistic form; hence, we give results only for that case.

It is easy to check that for one observation per cell if $t^* \equiv \Sigma_1^n x_j y_j$, t^* is a sufficient statistic for β (given vectors $x_1,\ldots,x_n$). Note that t^* is the sum of those x vectors for which a response was obtained ($y_j = 1$). In this case, the MLE of β, $\hat{\beta}$, must

satisfy the vectorial equation

$$\sum_{j=1}^{n} \left(1 + e^{-x_j'\hat{\beta}}\right)^{-1} x_j = t^* = \sum_{j=1}^{n} x_j y_j \ . \qquad (12)$$

Also, it is straightforward to check that log L is globally concave, so that (12) provides an absolute maximum.[7] The probability of a response at stimulus level x is then estimated by

$$\hat{p} = \left(1 + e^{-x'\hat{\beta}}\right)^{-1} \ .$$

---

[7] For a proof in the more general case of a polytomous dependent variable, see the discussion at the end of Section A.3.

A.3. ONE POLYTOMOUS QUALITATIVE VARIABLE

In many situations, the response variable is qualitative but there are more than two categories into which it may fall. Such categorical variables are called polytomous.

Let $y_1,\ldots,y_n$ denote independent (but not identically distributed) observations of categorical variables, and define

$$p^*_{ij} \equiv P\{y_i = a_j\} , \tag{13}$$

where the response variable for the ith subject, y_i, may assume Q values $a_1,\ldots,a_Q$; $i = 1,\ldots,N$; then $\sum_{j=1}^{Q} p^*_{ij} = 1$. We next relate the response probabilities to stimulus variables by means of the standardized multivariate logistic distribution cdf, which is defined by

$$F(t_1,\ldots,t_n) \equiv \frac{1}{1 + \sum_{j=1}^{n} e^{-t_j}} , \quad -\infty < t_j < \infty . \tag{14}$$

A symmetric form of (14) suggested by Mantel (1966) with the t_j appropriately defined in terms of the z_j and n related to Q, is

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{Q} e^{z_k}} , \qquad j = 1,\ldots,Q . \tag{15}$$

It will be seen in (24) below that this symmetric form is useful for characterizing several correlated polytomies (multidimensional contingency tables). When Q = 2, the model in (15) may be reduced to that of (7) by taking $p_1 \equiv p$, and $z_2 - z_1 \equiv x'\beta$.

The general polytomous model of (15) with the z_j expressed as linear functions of unknown parameters may be estimated by maximum likelihood, as was done in Section A.2 for the dichotomous model. For

simplicity of notation we permit the y_i to be multiple observations of the same group of objects; but we ignore such information, since it is not needed in the MLE approach.

Using (13), adopt the model

$$p^*_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{Q} e^{z_{ik}}}, \qquad i = 1,\ldots,N; \quad j = 1,\ldots,Q, \tag{16}$$

and take $z_{ij} \equiv x_i'\gamma_j$ for fixed x_i: s x 1, and γ_j a vector of coefficient weights. Let $Z(N \times Q) \equiv (z_{ij})$ denote that Z is an N x Q matrix with general element z_{ij}. Then if $Z(N \times Q) \equiv (z_{ij})$, $X(p \times N) \equiv (x_1,\ldots,x_N) = (x_{ij})$, $\Gamma(p \times Q) \equiv (\gamma_1,\ldots,\gamma_Q) = (\gamma_{ij})$, $Z = X'\Gamma$ and Γ must be estimated for given X. We take $\sum_1^Q \gamma_j = 0$ so that all coefficients will be uniquely defined. The model in (16) was studied by Bock (1970) and independently by Press (1972, pp. 268-272). Walker and Duncan (1967) proposed a weighted least squares solution, which is equivalent to a maximum likelihood estimation solution in large samples. The maximum likelihood estimation solution for Γ is given below.

Define $v_{ij} = 1$ if $y_i = a_j$ and let $v_{ij} = 0$ otherwise. The likelihood function is

$$L = \prod_{i=1}^{N} \prod_{j=1}^{Q} p^{*v_{ij}}_{ij}, \quad \sum_{j=1}^{Q} p^*_{ij} = 1, \quad \sum_{j=1}^{Q} v_{ij} = 1 .$$

Note that $t_j \equiv \sum_{i=1}^{N} x_i v_{ij}$ is sufficient for Γ given X. The result of maximizing L subject to $\sum_1^Q \gamma_j = 0$ is that the MLE of γ_ℓ, $\hat{\gamma}_\ell$, must satisfy the system of equations

$$\sum_{i=1}^{N} \left[\frac{e^{x_i'\hat{\gamma}_\ell}}{\sum_{k=1}^{Q} e^{x_i'\hat{\gamma}_k}} \right] x_i = \sum_{i=1}^{N} x_i v_{i\ell}, \tag{17}$$

$$\text{and} \quad \sum_{j=1}^{Q} \hat{\gamma}_j = 0; \; \ell = 1,\ldots,Q. \tag{18}$$

The solution to this system of equations must, of course, be obtained by numerical methods. Whether it is preferable to solve (17) and (18) or to maximize the original likelihood function depends upon the particular computer software available.

Note that (17) and (18) are only first order conditions for a maximum. The fact that they guarantee a maximum is seen from the second order conditions.

Let the log of the likelihood be given as

$$L^* \equiv \log L = \sum_{j=1}^{Q} t_j'\gamma_j - \sum_{i=1}^{N} \log \left[\sum_{k=1}^{Q} e^{x_i'\gamma_k} \right] .$$

It is straightforward to show that

$$\frac{\partial^2 L^*}{\partial \gamma_m \gamma_m'} = - \sum_{i=1}^{N} \delta_i^2 (m)(x_i x_i') ,$$

where

$$\delta_i^2(m) \quad \frac{\sum_{k=1}^{Q} \exp\{x_i'(\gamma_k - \gamma_m)\} - 1}{\left[\sum_{k=1}^{Q} \exp\{x_i'(\gamma_k - \gamma_m)\} \right]^2} .$$

It is clear by inspection that $\delta_i^2(m) \geq 0$, and since $(x_i x_i')$ is positive semi-definite for all i, L* is globally concave; so a true maximum corresponds to the first order conditions.

The results in (17) and (18) will form the basis for parameter estimation in the general, multi-equation, log-linear model (that is, the model that interrelates many categorical variables through other explanatory variables and interactions). The general model is described in Section A.4.

A.4. SEVERAL QUALITATIVE POLYTOMOUS VARIABLES

GENERAL MODEL

In earlier sections our discussion concerned a single qualitative variable, for which the probability that it takes on a specified value depends on certain exogenous or explanatory variables. In this section we explore the more general (multivariate) case of any number of jointly varying polytomous variables and exhibit the relation between the log-linear model of contingency table analysis and the standard logistic model described in Section A.3. After we define the basic model, we discuss estimation and hypothesis testing.

In this analysis it is not the *value* of a particular qualitative variable, be it dichotomous or polytomous, that is important but the *probability* that the variable takes on a particular value. That the probability distribution is the central concept in the usual econometric context is often obscured by the emphasis on structural and reduced form equations; yet, it is the joint probability distribution of the unlagged endogenous variables, and various conditional distributions (or certain endogenous variables given others), that are at the heart of the analysis.

Logit analysis of several related qualitative variables by weighted least squares was proposed by Grizzle, Starmer, and Koch (1969) and by Grizzle (1971). Although the results are applicable in large samples, they ignore the probability structure of the errors, and therefore they may not be too efficient in small or moderate samples. The model described below uses all of the available information as the basis for inference. It also uses a maximum likelihood estimation procedure, which may be regarded as desirable.

Consider the general *multivariate polytomy*, which is often referred to as a *multidimensional contingency table*. In this case we have q variables, $y_1,\ldots,y_q$, which may take on any one of $I_1,\ldots,I_q$ *unordered* alternative values. Whether the alternative values are ordered is quite important theoretically but perhaps of less significance

practically.[8] The total number of possibilities (cells) is

$$Q = \prod_{k=1}^{q} I_k \; ; \tag{19}$$

for example, $Q = 4$ in the case of two dichotomous variables. It can easily be seen that as long as the problem can be treated as one in which the margins are not fixed, or other restrictions introduced, the q-variate polytomy with polytomies of order $I_1,\ldots,I_q$ is equivalent to a univariate polytomy of order Q; let $p_{i_1,i_2,\ldots,i_q}$, with $i_1 = 1,\ldots,I_1; i_q = 1,\ldots,I_q$, be the probabilities attaching to the joint events that y_1 takes on the i_1th value, y_2 takes on the i_qth value, and so on; then redefine

$$\begin{aligned} p_1 &= p_{1,\ 1,\ldots,1}, \\ &\vdots \quad \vdots \\ p_Q &= p_{I_1,\ I_2,\ldots,I_q}, \end{aligned} \tag{20}$$

which correspond to the probabilities of the events that some polytomous variable z_k takes on the first, second,..., or Qth value. Arguing in the other direction, since any number can be written in binary form, any polytomy may be expressed in terms of a multivariate dichotomy. This equivalence is useful in the analysis of jointly dependent qualitative variables, for the probabilities for a set of q jointly dependent binary variables may always be re-expressed in terms of a single polytomy with 2^q possible values, and, conversely, any polytomy may be reformulated in terms of a joint occurrence of a certain sequence of binary variables.[9]

---

[8] Analysis of the ordered case may be quite complex; see Cox (1970, pp. 103-104). Mantel (1966, p. 91), however, suggests that the analysis not be constrained by the ordering: "But there is no need in applying the model to know the ordering. With adequate data the fitted ordering should follow the true parametric ordering. With inadequate data, inversions may occur, but the same kind of thing is true in ordinary regression analysis or probit analysis: one does not incorporate into the fitting the knowledge that the slope should be positive but depends rather on the adequacy of the data to bring this out."

[9] In general one will need some additional restrictions on the probabilities of the joint occurrence of these binary variables since the

Consider now the probabilities $p_{i_1,\ldots,i_q}$, and let $F(\cdot)$ denote a continuous monotone transformation with inverse $G(\cdot)$. Let

$$P_{i_1,\ldots,i_q} = F(p_{i_1,\ldots,i_q}) , \tag{21}$$

so that

$$p_{i_1,\ldots,i_q} = G(P_{i_1,\ldots,i_q}) , \tag{22}$$

for $i_1 = 1,\ldots,I_1;\ldots;i_q = 1,\ldots,I_q$. The function $F(\cdot)$ is taken to be the natural logarithm ($G(\cdot)$ is the exponential function) so that if the transformed cell probabilities are expressed in the linear form

$$P_{i_1,\ldots,i_q} = \mu + \theta_{i_1,\ldots,i_q}, \tag{23}$$

for $i_1 = 1,\ldots,I_1;\ldots;i_q = 1,\ldots,I_q$, the cell probabilities, whose q-dimensional summation must be unity, assume the representation

$$p_{i_1,\ldots,i_q} = \frac{e^{\theta_{i_1,\ldots,i_q}}}{\sum_{i_1=1}^{I_1} \cdots \sum_{i_q=1}^{I_q} e^{\theta_{i_1,\ldots,i_q}}} . \tag{24}$$

This is the symmetric form of the multivariate logistic function discussed in Section A.3. By choosing other functions for $F(\cdot)$ alternative models will result, such as the multivariate probit model (see Ashford and Sowden, 1970). In fact, renaming the θs:

---

polytomy will seldom be of order exactly a power of 2; hence, the next higher power of 2 must be used and the probabilities of occurrence of certain combinations set to zero. As a result, computational difficulties may arise equivalent to those that occur in contingency table analysis when some cells are empty. Consequently, in cases where the polytomy is not a power of 2, it is usually not preferable, for computational reasons, to transform it into a multivariate dichotomy. Moreover, such transformations may obliterate restrictions inherent in the problem--for example, those imposed when certain marginal totals are fixed as would be the case in stratified sampling.

$z_1 \equiv \theta_{1,\ldots,1}$; $\cdots$, $z_Q \equiv \theta_{I_1,\ldots,I_Q}$; and using (20), gives the univariate polytomy of (15),

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{Q} e^{z_k}} , \quad j = 1,\ldots,Q . \tag{15}$$

The parameters (or functions) z_k or $\theta_{i_1\ldots i_q}$, however, cannot be chosen fully independently because they are not identified fully by the parameters (or functions) p_k or $p_{i_1\ldots i_q}$, which are the basic parameters of the problem. To see this simply add a constant, c say, to each z_k in (15); clearly p_k is unaltered; thus some normalization rule must be imposed, such as $z_Q = 0$, or, more conveniently from a theoretical point of view (see also Section A.3 where the analogous argument is presented)

$$\sum_{k=1}^{Q} z_k = 0 , \tag{25}$$

or, equivalently,

$$\sum_{i_1=1}^{I_1} \cdots \sum_{i_q=1}^{I_q} \theta_{i_1\ldots i_q} = 0 . \tag{26}$$

The formulation of the log-linear model of contingency tables of Birch (1963), and developed further by Goodman (1968, 1969, 1970, 1971a, 1971b, 1972a, 1972b, 1972c) and others, introduced a convenient parameterization of the cell probabilities by decomposing the θs of (24) into main effects (functions of a single subscript i_k) and interaction effects (functions of two or more subscripts); the latter are, in turn, distinguished as bivariate interactions, trivariate interactions, and so forth. All of these are treated as constants by Birch, and later by Goodman (1968, 1969, 1970, 1971a, 1971b, 1972a, 1972b, 1972c),

although it is possible to generalize the analysis by allowing them to be functions of exogenous variables. If the main effects are denoted by the function

$$\alpha_k(i_k), \qquad k = 1,\ldots,q \ ,$$

all possible bivariate interactions are denoted by the $\binom{q}{2}$ functions

$$\beta_{jk}(i_j,i_k), \qquad j < k = 1,\ldots,q \ ,$$

and so on; then the log-linear model of (24) may be expressed in the form

$$\begin{aligned} \log p_{i_1,\ldots,i_q} - \mu = \theta_{i_1,\ldots,i_q} &= \alpha_1(i_1) + \ldots + \alpha_q(i_q) \\ &+ \beta_{12}(i_1,i_2) + \ldots + \beta_{q-1,q}(i_{q-1},i_q) \\ &+ \ldots \\ &+ \omega_{1,2,\ldots,q}(i_1,\ldots,i_q) \ , \end{aligned} \tag{27}$$

where the identifying constraints generally used are given by

$$\alpha_1(\cdot) = \alpha_2(\cdot) = \ldots = \alpha_q(\cdot) = 0 \ ,$$

$$\beta_{12}(i_1,\cdot) = \beta_{12}(\cdot, i_2) = \ldots = \beta_{q-1,q}(\cdot, i_q) = 0 \ ,$$

$$\vdots$$

$$\omega_{1,2,\ldots,q}(\cdot,i_2,\ldots,i_q) = \ldots = \omega_{1,2,\ldots,q}(i_1,\ldots,i_{q-1},\cdot) = 0 \ , \tag{28}$$

and

$$\mu \equiv - \log \sum_{i_1} \ldots \sum_{i_q} \exp(\theta_{i_1,\ldots,i_q}) \ .$$

A dot denotes summation over that index. It is easy to check that the conditions of (28) imply that of (26).

The model in (27) and (28) may be used to study the q jointly varying qualitative variables in terms of contributions to the response due to main effects of each variable, and in terms of interaction effects among variables, much as these variables could be studied using the conventional ANOVA (if the response variables were quantitative). The results in either model are nonspecific, however, in that typical conclusions from such an analysis are that there are, or are not, interactions, and that certain variables interact with certain other ones. If there are some additional explanatory (exogenous) variables available, a version of ANOCOVA or regression may be attempted.

Suppose we have readily available observations of a vector of exogenous variables, x (which might include a constant term), and for θ defined in (24) we take

$$\theta_{i_1,\ldots,i_q} \equiv x'\theta^*_{i_1,\ldots,i_q} , \qquad (29)$$

where θ^* is expanded in terms of main effects and interactions, as in (27). For example, $\alpha_1(i_1) \equiv x'\alpha^*_1(i_1)$, $\beta_{1z}(i_1,i_2) \equiv x'\beta^*_{12}(i_1,i_2)$, and so on. If x is in fact constant, we have the ANOVA type of log-linear model. If x contains *bona fide* explanatory variables, but no constant term, we have a logistic type of multivariate regression. If, however, x contains a "one" as one of its elements in addition to *bona fide* explanatory variables, an ANOCOVA type of log-linear model results. Some of the elements of θ^* might be taken to be zero, *a priori*, in any case, depending upon the available information surrounding the problem. Below we detail some useful properties of log-linear and logistic models and discuss problems of inference in these models.

PROPERTIES OF THE MULTIVARIATE LOGISTIC MODEL

It is clear that (14) and (15) are completely equivalent forms of the same distribution since (15) may be rewritten as

$$p_j = \frac{1}{1 + \sum_{\substack{k=1 \\ k \neq j}}^{Q} e^{-(z_j - z_k)}},$$

a special case of (14) with $n \equiv Q - 1$. Therefore, the properties of (15) or (24) may be studied by studying those of (14).

Theorem I

All marginal distributions associated with (14) are also logistic.

Proof: Let any subset of t_j in (14) become arbitrarily large and note that what remains is still logistic of some dimension.

Remark: A marginal distribution of (14) is a marginal of a continuous vector following a logistic cdf. A marginal distribution of (24), however, is a marginal of a discrete distribution. In particular, let $z \equiv (z_1,\ldots,z_Q)$ and $p(z) \equiv [p_1(z),\ldots,p_Q(z)]$, where $p_j(z)$ denotes the jth cell probability for a Q-dimensional multinomial distribution, and each cell probability is permitted to depend upon a vector z. Then, if $p_j(z)$ is defined as in (15), p(z) defines the cell probability vector of a standard multinomial distribution, and z indexes the distribution. The cell probabilities are each, in turn, related to the cdf of a multivariate logistic distribution in the elements of z. The marginals of (14) must be logistic, however, for the term multivariate logistic to have any conventional meaning.

Theorem II

All conditional distributions associated with (14) are also logistic.

Proof: Write (14) in terms of two subvectors of $t \equiv (t_j) = (t_1,\ldots,t_r; t_{r+1},\ldots,t_n)$, so that

$$F(t_1,\ldots,t_n) \equiv F(t) = \frac{1}{1 + \sum_{1}^{r} e^{-t_j} + \sum_{r+1}^{n} e^{-t_j}}.$$

Then, because

$$G_1(t_1,\ldots,t_r|t_{r+1},\ldots,t_n) \equiv \frac{F(t_1,\ldots,t_n)}{G_2(t_{r+1},\ldots,t_n)} ,$$

and from Theorem I, $c \equiv G_2(t_{r+1},\ldots,t_n) = \left(1 + \sum_{r+1}^{n} e^{-t_j}\right)^{-1}$,

$$G_1(t_1,\ldots,t_r|t_{r+1},\ldots,t_n) = \frac{1 + \sum_{r+1}^{n} e^{-t_j}}{1 + \sum_{r+1}^{n} e^{-t_j} + \sum_{1}^{r} e^{-t_j}} .$$

Equivalently,

$$G_1(t_1,\ldots,t_r|t_{r+1},\ldots,t_n) = \frac{1}{1 + c\sum_{1}^{r} e^{-t_j}} . \qquad (30)$$

Note that since $(t_{r+1},\ldots,t_n)$ is fixed, c is constant; moreover, (30) is a multivariate logistic cdf (non-standardized), which is standardized by subtracting log c from every t_j.

Theorem III

All marginal distributions associated with (24) are in multivariate logistic form but in variables different from those of the multivariate distribution.

Proof: Consider the univariate (discrete) marginal distribution of the distribution in (24):

$$r_{i_1} \equiv p_{i_1,\ldots\ldots} = \sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_q} p_{i_1\ldots i_q}$$

$$= \frac{\sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_q} e^{\theta_{i_1 \cdots i_q}}}{\sum_{i_1=1}^{I_1} \left\{ \sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_q} e^{\theta_{i_1 \cdots i_q}} \right\}}$$

$$= \frac{e^{H_1(i_1)}}{\sum_{i_1=1}^{I_1} e^{H_1(i_1)}},$$

where

$$H_1(i_1) = \log \sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_q} e^{\theta_{i_1 \cdots i_q}}.$$

To define the parameters uniquely, we impose the condition that $\sum_{i_j=1}^{I_j} H_j(i_j) = 0$ for every i, a condition that can be satisfied by the H_j as defined above. Clearly $p_{i_1,\ldots,}$ is in multivariate logistic form in $(H_1,\ldots,H_{I_1})$ and $p_{i_1,\ldots,i_q}$ is in multivariate logistic form in the θs. In the same way, the discrete multivariate marginal probabilities are also expressible in multivariate logistic form, for example,

$$r_{i_2,\ldots,i_q} \equiv \sum_{i_1} p_{i_1,\ldots,i_q} = \frac{\sum_{i_1=1}^{I_1} e^{\theta_{i_1 \cdots i_q}}}{\sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_q} \left\{ \sum_{i_1=1}^{I_1} e^{\theta_{i_1 \cdots i_q}} \right\}}$$

$$= \frac{e^{H_{2\ldots q}(i_2,\ldots,i_q)}}{\sum_{i_2=1}^{I_2} \cdots \sum_{i_q=1}^{I_2} e^{H_{2\ldots q}(i_2,\ldots,i_q)}},$$

where

$$H_{2\ldots q}(i_2,\ldots,i_q) = \log \sum_{i_1=1}^{I_1} e^{\theta_{i_1\ldots i_q}}.$$

For uniqueness of the parameters we require that

$$\sum_{i_2} \cdots \sum_{i_q} H_{2\ldots q}(i_2,\ldots,i_q) = 0.$$

Theorem IV

The conditional probabilities associated with (24) have a multivariate logistic representation and are given by

$$p_{\underline{i}|\bar{\underline{i}}} = \frac{\exp\left(\theta_{i_1,\ldots,i_q}\right)}{\sum_{\underline{i}} \exp\left(\theta_{i_1,\ldots,i_q}\right)},$$

where $\underline{i}$ denotes a subvector of $(i_1,\ldots,i_q)$, and $\bar{\underline{i}}$ denotes the subvector of $(i_1,\ldots,i_q)$ whose components are not included in $\underline{i}$.

Proof: Define the marginal distribution of the discrete variables corresponding to the elements of $\bar{\underline{i}}$, $r_{\bar{\underline{i}}} = \sum_{\underline{i}} p_{i_1,\ldots,i_q}$. Then, by definition,

$$p_{\underline{i}|\bar{\underline{i}}} = \frac{p_{i_1,\ldots,i_q}}{r_{\bar{\underline{i}}}} = \frac{\exp\left(\theta_{i_1,\ldots,i_q}\right) \Big/ \sum_{\underline{i}} \sum_{\bar{\underline{i}}} \exp\left(\theta_{i_1,\ldots,i_q}\right)}{\sum_{\underline{i}} \exp\left(\theta_{i_1,\ldots,i_q}\right) \Big/ \sum_{\underline{i}} \sum_{\bar{\underline{i}}} \exp\left(\theta_{i_1,\ldots,i_q}\right)},$$

which gives the required result.

Remarks: (1) If the parameters $\theta_{i_1,\ldots,i_q}$ are assumed to be functions of certain explanatory variables--for example, linear functions--it does not follow that the marginal probabilities will be logistic in functions of the same form (this result is implied in Theorem III). If the θs are linear functions, the Hs are highly nonlinear. Nonetheless, the conditional probabilities are logistic in functions of the same form.

(2) Theorem IV is used in the section on hypothesis testing to define the conditional odds ratio for the relative levels of one set of variables, given another set.

Theorem V

The qualitative random variables $y_1,\ldots,y_q$ defining the contingency table whose transformed cell probabilities have the representation given in (29) and (30) are mutually independent if and only if all interaction terms vanish.

Proof: This result may be shown directly.

If all interaction terms vanish, it is straightforward to check that the marginal probabilities are given by

$$r_{i_j} = \frac{e^{\alpha_j(i_j)}}{\sum_{i_k=1}^{I} e^{\alpha_k(i_k)}}, \qquad j = 1,\ldots,q .$$

But from (29), with no interactions,

$$p_{i_1,\ldots,i_q} = e^{\mu} \prod_{j=1}^{q} e^{\alpha_j(i_j)} = \prod_{j=1}^{q} r_{i_j} ;$$

so the y_j are independent.

Conversely, suppose the y_j are independent. Then,

$$p_{i_1,\ldots,i_q} \equiv \prod_{j=1}^{q} r_{i_j} .$$

where the $\equiv$ implies an identity in all i_j. Equivalently,

$$\log p_{i_1,\ldots,i_q} \equiv \sum_{j=1}^{q} \log r_{i_j} .$$

Substituting from (27) and Theorem III gives

$$\mu + \sum_{j=1}^{q} \alpha_j(i_j) + T(i_1,\ldots,i_q) \equiv q\mu + \sum_{j=1}^{q} H_j(i_j) ,$$

where $T(i_1,\ldots,i_q)$ denotes a sum of terms, each of which depends upon at least two of the i_j. For this relation to be an identity in all i_j, $T(\cdot)$ must be zero; that is, independence of $y_1,\ldots,y_q$ implies all the interactions must vanish. It is straightforward to check that under these conditions

$$\sum_{j=1}^{q} H_j(i_j) = \sum_{j=1}^{q} \alpha_j(i_j) + \mu(1 - q)$$

identically in all i_j. Note that

$$H_j(i_j) \neq \alpha_j(i_j) .$$

Since each $H_j(i_j)$ depends upon only a single i_j, $p_{i_1,\ldots,i_q}$ is expressible in general multivariate logistic form involving only main effects (all interaction terms are zero). Thus, under independence the model may be written

$$\log p_{i_1,\ldots,i_q} = \mu + \sum_{j=1}^{q} \alpha_j(i_j) .$$

Double Dichotomy

The results of this section may be illustrated for the bivariate case in which each qualitative variable is binary--that is, the double dichotomy. If y_1 and y_2 can each assume values 1 and 2, the joint probabilities are as depicted in the cells of the following contingency table.

y_1 \ y_2	1	2	
1	p_{11}	p_{12}	$p_{1.}$
2	p_{21}	p_{22}	$p_{2.}$
	$p_{.1}$	$p_{.2}$	1

The marginal probabilities are shown at the right and below. The parameters of the log-linear model are related to the cell probabilities through (27) and (28), which, in this case, if we take $p_{1,2}(i_1,i_2) \equiv p_{i_1,i_2}$, reduce to

$$\log p_{i_1,i_2} = \mu + \alpha_1(i_1) + \alpha_2(i_2) + \beta_{12}(i_1,i_2) ,$$

with the constraints that $\alpha_1(\cdot) = \alpha_2(\cdot) = 0$, and $\beta_{12}(\cdot,i_2) = \beta_{12}(i_1,\cdot) = 0$. Defining

$$\alpha_1 \equiv \alpha_1(1), \ \alpha_2 \equiv \alpha_2(1), \ \beta \equiv \beta_{12}(1,1)$$

for simplicity yields the following equivalent form:[10]

$$\begin{aligned} \log p_{11} &= \mu + \alpha_1 + \alpha_2 + \beta , \\ \log p_{12} &= \mu + \alpha_1 - \alpha_2 - \beta , \\ \log p_{21} &= \mu - \alpha_1 + \alpha_2 - \beta , \\ \log p_{22} &= \mu - \alpha_1 - \alpha_2 + \beta , \end{aligned} \tag{31}$$

[10] A generalization of this result to q dichotomous variables is given in Eq. (41).

and

$$-\mu = \log[\exp(\alpha_1 + \alpha_2 + \beta) + \exp(\alpha_1 - \alpha_2 - \beta)$$

$$+ \exp(-\alpha_1 + \alpha_2 - \beta) + \exp(-\alpha_1 - \alpha_2 + \beta)] \ . \qquad (32)$$

Thus, the double dichotomy may be parameterized in terms of $(\alpha_1,\alpha_2,\beta)$, or equivalently, in terms of, say (p_{11},p_{12}, p_{21}), with p_{22} being determined from $p_{11} + p_{12} + p_{21} + p_{22} = 1$.

It is well known, and straightforward to check, that the condition for independence in a double dichotomy, namely $p_{ij} = p_{i\cdot}p_{\cdot j}$, for $i = 1,2$ and $j = 1,2$, is satisfied if and only if the cross products of the cell probabilities are equal--that is, $p_{11}p_{22} = p_{12}p_{21}$. Checking the cross product condition in (31) shows that it is satisfied if and only if there is no interaction term in the log-linear model (see Theorem III).

Now define

$$z_1 \equiv \alpha_1 + \alpha_2 + \beta, \qquad z_2 \equiv \alpha_1 - \alpha_2 - \beta \ ,$$

$$z_3 \equiv -\alpha_1 + \alpha_2 - \beta, \qquad z_4 = -\alpha_1 - \alpha_2 + \beta \ .$$

Then from (31) and (32), if $p_1 \equiv p_{11}$, $p_2 \equiv p_{12}$, $p_3 \equiv p_{21}$, $p_4 \equiv p_{22}$, the log-linear model of (31) and (32) is cast into the form of the logistic model

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{4} z_k} \ , \qquad j = 1,\ldots,4 \ ,$$

so that the bivariate dichotomy now corresponds to a univariate polytomy (see (15)) with four unordered categories.

The conditional probabilities promised in Theorem IV are illustrated for the double dichotomy by

$$p\{y_2 = 1|y_1 = 1\} = \frac{p_{12}(1,1)}{r_1} = \frac{p_{11}}{p_{1\cdot}} = \frac{p_1}{p_1 + p_2}$$

$$= \frac{e^{z_1}}{e^{z_1} + e^{z_2}} ,$$

which is in the symmetric logistic form. Equivalently,

$$p\{y_2 = 1|y_1 = 1\} = \frac{1}{1 + e^{-2(\alpha_2+\beta)}} ,$$

which is in the classical (but unstandardized) logistic cdf form in $(\alpha_2 + \beta)$.

Suppose there are q dichotomous variables, $y_1,\ldots,y_q$ whose states are designated by k_1 and k_2. Then if

$$p\{y_1 = k_1|y_2 = y_3 \ldots = y_q = k_1\} \equiv p_{1|1,\ldots,1} ,$$

$$p_{1|1,\ldots,1} = \frac{1}{1 + e^{-2t_{1,\ldots,1}}} ,$$

where

$$t_{1,\ldots,1} \equiv \alpha_1(k_1) + \sum_{j=2}^{q} \beta_{1,j}(k_1, k_1) + \sum_{j=3}^{q} \gamma_{1,2,j}(k_1, k_1, k_1)$$

$$+ \ldots + \omega_{1,2,\ldots,q}(k_1,\ldots,k_1) ,$$

and the α, β, $\gamma,\ldots,\omega$ are defined in (27). This result follows immediately from Theorem IV when the $\theta_{i_1,\ldots,i_q}$ are defined as in (27).

To understand the relationships among conditioning variables in a double dichotomy, log-linear or logistic model, consider first the more general case of q qualitative variables.

Let $y \equiv (y_2,\ldots,y_q)$ and $j \equiv (j_2,\ldots,j_q)$, so that

$$P\{y_1 = i|y = j\} = \frac{p_{1,\ldots,q}(i,j)}{p_{1,\ldots,q}(\cdot,j)} .$$

Recall that a dot denotes a summation over an index. Then, the "odds ratio" is given by

$$\frac{P\{y_1 = i_1|y = j\}}{P\{y_1 = i_2|y = j\}} = \frac{p_{1,\ldots,q}(i_1,j)}{p_{1,\ldots,q}(i_2,j)} .$$

Since $\log p_{1,\ldots,q}(i_1,j) \equiv \mu + \theta_{i_1,j}$, the log of the odds ratio is

$$R_{i_1,i_2}(j) \equiv \log \frac{P\{y_1 = i_1|y = j\}}{P\{y_1 = i_2|y = j\}} = \theta_{i_1,j} - \theta_{i_2,j} . \qquad (33)$$

The relation in (33) shows that the log-odds is always a linear function of the θs. Thus, for $q = 2$, since

$$\theta_{i_1,j} = \alpha_1(i_1) + \alpha_2(j) + \beta_{12}(i_1,j) ,$$

substitution in (33) gives

$$R_{i_1,i_2}(j) = [\alpha_1(i_1) - \alpha_1(i_2)] + [\beta_{12}(i_1,j) - \beta_{12}(i_2,j)] .$$

If each variable is binary, as in the double dichotomy example above,

$$\alpha_1(i_1) + \alpha_1(i_2) = 0, \qquad \beta_{12}(i_1,j) + \beta_{12}(i_2,j) = 0 ,$$

so that

$$R_{i_1,i_2}(j) = 2[\alpha_1(i_1) + \beta_{12}(i_1,j)] \ .$$

Thus, if y_1 and y_2 are independent, $\beta_{12}(i_1,j) = 0$ and $R_{i_1,i_2}(j)$ depends only on the main effect for y_1. In the absence of independence the log-odds depends on the levels of the conditional variables only through the interaction term.

The result in (33) provides a means of analyzing simultaneously determined qualitative variables, in particular simultaneously determined dichotomous variables, in terms of the analogue of the structural equations in standard econometric analysis. This analogue is just the conditional probability function.

In the more general multivariate case, the conditional probabilities depend only on own main effects and on those interaction effects that involve interactions between the conditioned and conditioning variables or between conditioning variables. This is seen explicitly below. From Theorem IV, the conditional probability of any subset of y_j given the levels of the remaining y_j is expressible as

$$p_{\underline{i}|\overline{\underline{i}}} = \frac{\exp\ (\theta_{i_1,\ldots,i_q})}{\sum\limits_{\underline{i}} \exp\ (\theta_{i_1,\ldots,i_q})} \ .$$

Now write the $\theta_{i_1,\ldots,i_q}$, using (27), in the form

$$\theta_{i_1,\ldots,i_q} = \phi_{\underline{i}} + \phi_{\underline{\underline{i}}} \ ,$$

where

$$\phi_{\underline{i}} \equiv \sum_{j\epsilon\underline{i}} \alpha_j(i_j) + \text{(interactions involving elements of } \underline{i}),$$

$$\phi_{\underline{\underline{i}}} \equiv \sum\nolimits_{j\epsilon\underline{i}} \alpha_j(i_j) + \text{(interactions not involving elements of } \underline{i}).$$

Substituting for $\theta_{i_1,\ldots,i_q}$ shows that $\phi_{\underline{i}}$ drops out and the result depends only on main effects in $\underline{i}$, and on interactions between the two sets of variates--that is,

$$P_{\underline{i}|\bar{\underline{i}}} = \frac{\exp(\phi_{\underline{i}})}{\sum_{\underline{i}} \exp(\phi_{\underline{i}})} . \tag{34}$$

Note that $\phi_{\underline{i}}$ may involve both interactions between variables in $\underline{i}$ and in $\bar{\underline{i}}$, and interactions between variables in $\underline{i}$. Interactions between variables only in $\bar{\underline{i}}$ are excluded, however.

It is sometimes both convenient and appropriate to ignore interaction effects of order higher than two and to assume that main effects are functions of the exogenous variables but interaction effects are not. Then, from Theorem IV, it is possible to express the conditional odds ratio for each endogenous variable as a linear function of the exogenous variables and those interaction terms involving the endogenous variable in question with the others, which are treated as conditioning (see Eq. (34)). Moreover, when all endogenous variables are binary, symmetric forms are obtainable by treating the endogenous binary variables as +1 or -1 (see Eq. (41)). The choice of the multivariate logistic distribution in the above models may be given a justification in terms of a physical mechanism by extending the argument on p. 14.

Let $E_1,\ldots,E_Q$ denote a collection of mutually exclusive and exhaustive events (such as the event of falling into each of the Q cells of a multi-dimensional contingency table) and let x denote a p-vector with continuous density of $h(x|\Theta)$, for some parameter matrix Θ. By Bayes theorem,

$$P\{E_k|x\} = \frac{P\{E_k\}h(x|E_k,\Theta)}{\sum_{j=1}^{Q} P\{E_j\}h(x|E_j,\Theta)} .$$

Let $q_j \equiv P\{E_j\}$ and let $\mathcal{L}(x) = N(\theta_j,\Sigma)$ if E_j is true, $j = 1,\ldots,Q$.

Then,

$$P\{E_k|x\} = \frac{q_k e^{-1/2(\theta_k'\Sigma^{-1}\theta_k - 2\theta_k'\Sigma^{-1}x)}}{\sum_{j=1}^{Q} q_j e^{-1/2(\theta_j'\Sigma^{-1}\theta_j - 2\theta_j'\Sigma^{-1}x)}}$$

$$\equiv \frac{e^{a_k + b_k'x}}{\sum_{j=1}^{Q} e^{a_j + b_j'x}} , \qquad (35)$$

for an obvious set of weights $\{a_j, b_j\}$. Since renaming the θs in (24) and making them linear functions of some exogenous vector x yields (35), it is seen that the basic logistic transformation in (24) could have arisen in a fundamental way.

INFERENCE IN THE MULTIVARIATE LOG-LINEAR/LOGISTIC MODEL

The multivariate log-linear/logistic models discussed earlier in this section have been considered so far only from a probabilistic, model formulation viewpoint. We now consider how to fit the models to real data and how to study alternative hypotheses about the models on the basis of observed evidence.

Estimation

It is well known (and straightforward to check) that if $p_{1,2,\ldots,q}(i_1,\ldots,i_q)$ denotes the probability of falling in cell $(i_1,\ldots,i_q)$ of a q-dimensional contingency table, and if $N_{i_1,\ldots,i_q}$ denotes the number of observations falling into cell $(i_1,\ldots,i_q)$, the MLEs of the cell probabilities are the sample cell frequencies; that is, if $\hat{p}$ denotes the MLE,

$$\hat{p}_{1,2,\ldots,q}(i_1,\ldots,i_q) = \frac{N_{i_1,\ldots,i_q}}{N} , \qquad (36)$$

where $N \equiv \sum_{i_1} \cdots \sum_{i_q} N_{i_1,\ldots,i_q}$ denotes the total number of observations in the contingency table.

Now suppose the cell probabilities are expressed as exponential functions of the main effects and interaction effects according to the log-linear model. The parameters of the log-linear model may be estimated by ML by solving for the parameters in terms of the cell probabilities and substituting the MLEs of the cell probabilities.

In the case of the double dichotomy discussed above (p. 77), the parameter relationships are as given in (31). In that case, for illustrative purposes, (31) may be solved simultaneously to obtain

$$\alpha_1 = (1/4) \log \frac{p_{11}\, p_{12}}{p_{12}\, p_{22}}, \qquad \alpha_2 = (1/4) \log \frac{p_{11}\, p_{21}}{p_{12}\, p_{22}},$$

and

$$\beta = (1/4) \log \frac{p_{11}\, p_{22}}{p_{12}\, p_{21}}. \tag{37}$$

Therefore, the MLEs are given by

$$\hat{\alpha}_1 = (1/4) \log \frac{N_{11}\, N_{12}}{N_{21}\, N_{22}}, \qquad \hat{\alpha}_2 = (1/4) \log \frac{N_{11}\, N_{21}}{N_{12}\, N_{22}},$$

and

$$\hat{\beta} = (1/4) \log \frac{N_{11}\, N_{22}}{N_{12}\, N_{21}}. \tag{38}$$

It will always be possible to solve for the MLEs of the main effects and interactions in any q-dimensional polytomy in the same way. All relationships will always be linear in the logs of the cell

probabilities, so in every case[11] the result will be completely analogous to (38). This situation, where all parameters of the log-linear model are present, is the one Goodman calls the "saturated" model (see Goodman, 1970, p. 253). In some situations, however, it may be desirable to set some of the parameters equal to zero, *a priori*, and then estimate the remaining parameters of this "unsaturated" model by MLE. Unfortunately, except for some simple special cases, estimation of the unsaturated model is generally iterative and therefore much more difficult computationally (see Bishop, 1969, for a summary of various suggested procedures). For example, it is necessary to estimate an unsaturated model when one desires to estimate the parameters of a log-linear model in which the qualitative variables are mutually independent (so that by Theorem III, all interaction parameters must be taken to be zero); or, for example, when interactions of order higher than two are to be ignored. Estimation of parameters in a contingency table, under restricted conditions by MLE, was begun by Bartlett (1935) (as suggested by R. A. Fisher) for the case of a 2 x 2 x 2 table with fixed margins. His work was extended by Roy and Kastenbaum (1956) to the case of an r x s x t table with no three-factor interaction term. The work of Birch (1963) showed that marginal totals are MLEs of their expectations (and are sufficient statistics for these parameters). The "iterative scaling" technique described by Bishop (1969) and used extensively by Goodman (Goodman, 1970, p. 237) provides a rapid computational algorithm for estimating the parameters of unsaturated log-linear models of the ANOVA type.

If the log-linear/logistic model is of the regression type, ML may be used directly to estimate the coefficients of the explanatory variables, in addition to the effects themselves. The ANOVA parameterization is a special case of the general model and can be studied by

[11] Difficulties arise in the event of empty cells (see, for example, Fienberg, 1972, for a discussion of this case). In the unsaturated log-linear model, it may be possible to identify some of the parameters even when some of the cells are empty.

placing the ANOVA type of log-linear model into the regression type of log-linear model formulation.

Suppose a q-variate polytomy with each of the qualitative variables related to a p-vector of exogenous variables, x. Let $p^*_{ij} \equiv P\{\text{ith subject falls into cell } j\}$, $i = 1,\ldots,N$, and $j = 1,\ldots,Q$, with $Q = \prod_{k=1}^{q} I_k$ from (19). The logistic model gives

$$p^*_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{Q} e^{z_{ik}}}, \qquad i = 1,\ldots,N; \quad j = 1,\ldots,Q . \tag{16}$$

Now with $z_{ij} \equiv x_i'\gamma_j$, for some observed p-vectors of exogenous variables, $x_1,\ldots,x_N$, the MLEs of the γ_j are found as the solutions of (17) and (18), as in Section A.3.

To fit the logistic model in the manner shown above requires only the logistic transformation assumption. However, if we impose the ANOVA type of structure of (27), the terms in the exponents in (16) will be expressed as main effects and interaction terms, each of which may be expressed as a linear function of exogenous variables (see the formulation in (29)). To facilitate understanding consider the double dichotomy example (p. 77). The model may be parameterized as in (31) and (32). Now suppose

$$\alpha_1 = x(1)'\gamma(1), \quad \alpha_2 = x(2)'\gamma(2), \quad \text{and } \beta = x(3)'\gamma(3), \tag{39}$$

where x(1), x(2), x(3) are each vectors of exogenous variables, and $\gamma(1)$, $\gamma(2)$, and $\gamma(3)$ are vectors of weights. Then, for example,

$$\log p_{11} = \mu + x(1)'\gamma(1) + x(2)'\gamma(2) + x(3)'\gamma(3) ,$$

or

$$\log p_{11} = \mu + x'\gamma , \tag{40}$$

where $x' \equiv (x(1)', x(2)', x(3)')$, and $\gamma' = (\gamma(1)', \gamma(2)', \gamma(3)')$. When (39) is expressed with appropriate subscripts for observed data, it is equivalent to (16) with $z_{ij} = x_i'\gamma_j$. It is merely a matter of setting some of the elements of the γ vector equal to zero *a priori* to complete the comparison.[12] The advantage of the latter formulation is that specific main effects or interaction terms may be "explained" in terms of exogenous variables believed to be appropriate, *a priori*. In particular, if x_i is the value of x for the ith subject, and γ_1 is the coefficient vector corresponding to cell (1,1),

$$\log p^*_{i1} = \mu + x_i'\gamma_1 = \mu + \sum_{j=1}^{3} x_i(j)'\gamma_1(j), \qquad i = 1,\ldots,N ,$$

in exact agreement with (16).

The approach used above for the double dichotomy may of course be generalized to any multidimensional model. In particular, the approach is used for the multivariate case with only bivariate non-zero interaction effects in a computer program developed especially for implementation of these results.

In situations where the main effects and interactions have been estimated and it is desired to estimate the cell probabilities, the latter may easily be reconstructed from the former by substituting the estimated parameters into the right hand side of (27). In problems involving endogenous variables that are all dichotomous, it is sometimes convenient to denote the states by one and minus one and then define all main effects and interactions in terms of the values in the "one" states. Suppose the cell probabilities are defined as in (27), subject to the constraints in (28), and

$$\alpha_1 \equiv \alpha_1(1),\ \alpha_2 \equiv \alpha_2(1),\ldots,\alpha_q \equiv \alpha_q(1) \ ;$$

$$\beta_{1,2} \equiv \beta_{1,2}(1),\ \beta_{1,3} \equiv \beta_{1,3}(1),\ldots,\beta_{q-1,q} \equiv \beta_{q-1,q}(1);\ \ldots;$$

[12] Elements of γ could be set equal to any pre-assigned values other than zero as well, which might be useful in some circumstances.

$$\omega_{1,2,\ldots,q} \equiv \omega_{1,2,\ldots,q}(1,\ldots,1) \ .$$

Then, it is straightforward to check that any cell probability may be expressed as

$$\begin{aligned} \log p_{i_1,\ldots,i_q} = \mu &+ u_1\alpha_1 + u_2\alpha_2 + \ldots + u_q\alpha_q \\ &+ u_1u_2\beta_{1,2} + u_1u_3\beta_{1,3} + \ldots + u_{q-1}u_q\beta_{q-1,q} \\ &+ \ldots \\ &+ u_1u_2 \ldots u_q\omega_{1,2,\ldots,q} \ , \end{aligned} \tag{41}$$

where $u_j = +1, -1$ for all $j = 1,\ldots,q$, depending on the state expressed by u_j.

Hypothesis Testing

A likelihood-ratio test may be used to test any hypotheses about the model, for example, the hypothesis of independence. In general, the likelihood ratio, λ, is the ratio of the value of the likelihood function maximized under whatever constraints are embodied in the hypothesis being tested to the value maximized under no constraints except, of course, those implicit in the general model. Although in small samples the distribution of λ is very complicated, in large samples the quantity $-2 \log \lambda$ is distributed as Chi square with as many degrees of freedom as there are independent restrictions embodied in the hypothesis being tested, relative to its alternative. Hypothesis tests of relationships in a contingency table are not new and can be traced back to the work of Karl Pearson (1900). It was not until the recent work on the log-linear model of Goodman (1969) for three-way tables, and his extensions to m-way tables (in Goodman, 1970), that satisfactory hierarchical methods were proposed for partitioning Chi square

statistics based upon the likelihood ratio criterion. Goodman suggests we might start by testing that the highest order interactions in a log-linear model are zero. If the hypothesis is accepted, we might test that the next lower order interaction terms are zero given that the higher order interaction terms are zero, and so on. Proceeding in this way, the likelihood ratio test statistics, λ, are independent, and asymptotically, $-2 \log \lambda$ is distributed as Chi square.

"Causal relationships" and associations among the qualitative variables in a contingency table are studied using the log-linear model by estimating the parameters of the log of the odds ratio for a given qualitative variable expressed as a linear function of the effects, as in (27). The details of this type of approach may be found in Bishop (1969), and Goodman (1971a, 1972a, 1972b, 1972c). This type of analysis is appropriate when the only information available is a cross tabulation of qualitative variables. However, if there are some continuous, or discrete but ordered, exogenous variables available that can be related to individual subjects, they should be introduced where appropriate, and hypotheses concerning their coefficients may be tested by likelihood ratio procedures.

Testing whether a particular exogenous variable or even a group of such variables is a significant determinant of the probabilities can be carried out by the following likelihood ratio test: Replace the parameter values by zero; maximize subject to any additional constraints, and compare the result with the result unconstrained except by the maintained constraints of the model; -2 times the difference is distributed asymptotically as Chi square with as many degrees of freedom as coefficients set to zero.

To carry out tests using the parameterization of (15), with $z_j \equiv x'\gamma_j$, $j = 1,\ldots,Q$, we will need to consider linear functions of the γ_j. For example, in the double dichotomy, independence of the two qualitative variables is implied by $p_{11}\, p_{22} = p_{12}\, p_{21}$, or equivalently, by $\theta \equiv \gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 = 0$. Whether or not $\theta = 0$ may be studied by

studying the distribution of $\hat{\gamma}_1 - \hat{\gamma}_2 - \hat{\gamma}_3 + \hat{\gamma}_4$, which is normal in large samples since $\hat{\gamma}_j$ is a MLE. The asymptotic variances and covariances of the $\hat{\gamma}_j$ are easily obtained from the inverse of the information matrix based upon the likelihood parameterized in terms of the γ_j.

More generally, we may want to test hypotheses of the form:

$$H: \quad L\Gamma M = 0 \ ,$$

where $\Gamma \equiv (\gamma_1,\ldots,\gamma_Q)$: p x Q is the matrix of model parameters and L and M are arbitrary preassigned matrices. Likelihood ratio tests are still applicable and $L\hat{\Gamma}M$ is normally distributed in large samples.

Hypotheses may be tested using the original model or in terms of a conditional probability model involving only the parameters of interest. Since certain endogenous variables or levels of these variables are held fixed in the conditional probability model (and are therefore treated as exogenous), estimators based on this model will not be the same as those obtained from the unconstrained (unconditioned, jointly estimated) model. We term the estimators obtained from the conditional probability model *conditional* estimators, and those from the unconstrained model *full-information* estimators.

CONDITIONAL PROBABILITIES AND "CONDITIONAL ESTIMATORS"

In many situations we are concerned with the estimation of *conditional* probabilities and tests of hypotheses about those conditional probabilities. In the context of jointly dependent qualitative variables, the conditional probability function corresponds to the *structural equation* of the simultaneous equations approach (of traditional econometrics) to the analysis of continuously variable jointly dependent endogenous variables. For example, in the analysis of prices and quantities jointly determined by supply and demand, our interest frequently centers on the elasticity of demand, which expresses the

variation of the quantity demanded *conditional* on variations in price.[13] This section explores more fully the form of the conditional probability functions for several jointly dependent dichotomous variables; in the process, we find an interpretation of what econometricians have frequently done when confronted with problems involving multiple qualitative endogenous variables.

Consider, for example, the trivariate dichotomous case of the log-linear model, (27) and (28) above. (The main and interaction effects may be functions of explanatory variables, but such dependencies are here suppressed.) Let the indices i_1, i_2, i_3 take on the convenient values 1 for occurrence, and 0 for nonoccurrence. In the trivariate dichotomy, there are 8 - 1 = 7 distinct probabilities corresponding to the three main effects:

$$\alpha_1, \quad \alpha_2, \alpha_3 \quad ;$$

the three bivariate interaction effects:

$$\beta_{12}, \beta_{13}, \beta_{23} \quad ;$$

and the one trivariate interaction effect:

[13] Henry Schultz (1938) and others of his day did indeed attempt to estimate demand equations directly without reference, or with only limited reference, to the simultaneous systems in which these relations are embedded. The development of full and limited information maximum likelihood methods of estimation in the 1940s permitted a more sophisticated approach, corresponding to estimation based on the joint probability function of the jointly dependent endogenous variables. Here we suggest an analogue to what Schultz and others did in the case of demand functions for the estimation of conditional probability functions for jointly dependent qualitative variables. Although such methods are known to be inappropriate in the case of continuously variable data, they are common in what few analyses of qualitative data econometricians have undertaken to date.

γ_{123} .

Given the restrictions (28), all of the joint probabilities may be expressed in terms of these seven parameters (see also Eq. (41)):

$$\left.\begin{aligned}
\log p_{111} - \mu &= \alpha_1 + \alpha_2 + \alpha_3 + \beta_{12} + \beta_{13} + \beta_{23} + \gamma_{123} \\
\log p_{110} - \mu &= \alpha_1 + \alpha_2 - \alpha_3 + \beta_{12} - \beta_{13} - \beta_{23} - \gamma_{123} \\
\log p_{100} - \mu &= \alpha_1 - \alpha_2 - \alpha_3 - \beta_{12} - \beta_{13} + \beta_{23} + \gamma_{123} \\
\log p_{000} - \mu &= -\alpha_1 - \alpha_2 - \alpha_3 + \beta_{12} + \beta_{13} + \beta_{23} - \gamma_{123} \\
\log p_{001} - \mu &= -\alpha_1 - \alpha_2 + \alpha_3 + \beta_{12} - \beta_{13} - \beta_{23} + \gamma_{123} \\
\log p_{011} - \mu &= -\alpha_1 + \alpha_2 + \alpha_3 - \beta_{12} - \beta_{13} + \beta_{23} - \gamma_{123} \\
\log p_{101} - \mu &= \alpha_1 - \alpha_2 + \alpha_3 - \beta_{12} + \beta_{13} - \beta_{23} - \gamma_{123} \\
\log p_{010} - \mu &= \alpha_1 + \alpha_2 - \alpha_3 - \beta_{12} + \beta_{13} - \beta_{23} + \gamma_{123}
\end{aligned}\right\} \quad (42)$$

The parameter μ is defined so that

$$\sum_{i_1,i_2,i_3=0}^{1} p_{i_1 i_2 i_3} = 1 .$$

Using the result of Theorem IV the conditional probabilities may also be expressed in an instructive way. Take, for example, the probability of the occurrence of the first event conditional on the second and third events (see Eq. (34)):

$$\left.\begin{aligned}
\log p_{i_1=1}(i_2=1,i_3=1) - \mu_i^* &= \alpha_1 + \beta_{12} + \beta_{13} + \gamma_{123} \\
\log p_{i=1}(i_2=1,i_3=0) - \mu_1^* &= \alpha_1 + \beta_{12} - \beta_{13} - \gamma_{123} \\
\log p_{i_1=1}(i_2=0,i_3=1) - \mu_1^* &= \alpha_1 - \beta_{12} + \beta_{13} - \gamma_{123} \\
\log p_{i_1=1}(i_2=0,i_3=0) - \mu_1^* &= \alpha_1 - \beta_{12} - \beta_{13} + \gamma_{123}.
\end{aligned}\right\} \quad (43)$$

The parameter μ_1^* is given by

$$\mu_1^* \equiv - \log \left\{ \exp(\alpha_1 + \beta_{1,2} + \gamma_{1,2,3}) + \exp(-1)(\alpha_1 + \beta_{1,2} + \beta_{1,3} + \gamma_{1,2,3}) \right\}$$

The conditional probabilities $p_{i_2=1}(i_1,i_3)$ and $p_{i_3=1}(i_1,i_2)$ may be defined similarly.

Equations (43) may be rewritten in an interesting and useful form by introducing new variables u_1, u_2, u_3, which take on the values of +1 or -1 depending on the occurrence of the first, second, or third event (see Eq. (41)). (Earlier, we introduced the variables y_1, y_2, y_3, taking on the values 1 or 0; the u represent a simple rescaling of such variables more convenient for our present purpose.)[14] Equations (43) may be re-expressed in terms of a univariate logistic formulation:

$$\left.\begin{aligned}
p_{i_1=1}(i_2=1,i_3=1) &= \frac{e^{\{\alpha_1+\beta_{12}+\beta_{13}+\gamma_{123}\}}}{e^{\{\alpha_1+\beta_{12}+\beta_{13}+\gamma_{123}\}} + e^{-\{\alpha_1+\beta_{12}+\beta_{13}+\gamma_{123}\}}} \\
&= \frac{1}{1 + e^{-2\{\alpha_1+\beta_{12}+\beta_{13}+\gamma_{123}\}}} , \\
p_{i_1=1}(i_2=1,i_3=0) &= \frac{1}{1 + e^{-2\{\alpha_1+\beta_{12}-\beta_{13}-\gamma_{123}\}}} , \\
p_{i_1=1}(i_2=0,i_3=1) &= \frac{1}{1 + e^{-2\{\alpha_1-\beta_{12}+\beta_{13}-\gamma_{123}\}}} , \\
p_{i_1=1}(i_2=0,i_3=0) &= \frac{1}{1 + e^{-2\{\alpha_1-\beta_{12}-\beta_{13}+\gamma_{123}\}}} ,
\end{aligned}\right\} \quad (44)$$

so that, in compact form,

[14] The scaling of a qualitative variable is completely arbitrary in a theoretical sense. See p. 88 for a previous use of their scaling device.

$$P_{i_1=1}(i_2,i_3) = \frac{1}{1 + e^{-2\{\alpha_1+\beta_{12}u_2+\beta_{13}u_3+\gamma_{123}u_2u_3\}}} \cdot \tag{45}$$

Equation (45) suggests an analogue to the ordinary least squares estimation of individual structural equations within a system of simultaneous economic relationships. We know that such estimates have undesirable statistical properties and are inappropriate in the sense that they do not reflect the simultaneity of the system in which the individual structural relation is embedded. With each value of y_{1n}, $n = 1,\ldots,N$, reflecting the occurrence (1) or nonoccurrence (0) of the first event, are values of the variables u_{2n} and u_{3n}, reflecting the associated occurrence (+1) or nonoccurrence (-1) of the second and third events. The "likelihood function" associated with the sequence of observations (y_{1n}, u_{2n}, u_{3n}), $n = 1,\ldots,N$, treating u_{2n} and u_{3n} *as if they were exogenous*, is:

$$L^*(\alpha_1,\beta_{12},\beta_{13},\gamma_{123}) =$$

$$\prod_{n=1}^{N} \left[\frac{e^{t_n}}{e^{t_n} + e^{-t_n}}\right]^{y_{1n}} \left[\frac{e^{-t_n}}{e^{t_n} + e^{-t_n}}\right]^{1-y_{1n}} \tag{46}$$

where

$$t_n \equiv \alpha_1 + \beta_{12}\, u_{2n} + \beta_{13}\, u_{3n} + \gamma_{123}\, u_{2n}\, u_{3n} \cdot$$

Maximizing L* with respect to the parameters α_1, β_{12}, β_{13}, γ_{123}, which appear in the conditional likelihood function for the first event, yields an interesting kind of estimator of these parameters (to be consistent the *endogenous qualitative variables y_2 and y_3 should be rescaled as u_2 and u_3, to take on the values +1 or -1.* Such estimators are *not* appropriate estimators if one really believes the trivariate dichotomous model for the qualitative variables y_1, y_2, y_3, but just as ordinary least squares estimates of the parameters in a

structural equation may not be far off the mark compared with full information maximum likelihood estimators, so the above estimators may not be "bad" estimators of α_1, β_{12}, β_{13} and γ_{123}. Moreover, they may be a great deal cheaper to compute. We call estimators based on maximization of L* *conditional estimators*. Application of this method to each of the three conditional probability functions (for example, by using a univariate dichotomous program), yields one estimate each of α_1, α_2, and α_3, but *two* estimates each of β_{12}, β_{13}, and β_{23}, and *three* distinct estimates of γ_{123}.

In the computer program developed for this problem, we restrict our model to a simple and convenient form by assuming that: (a) interaction effects of order higher than two are absent; (b) bivariate interaction effects are constant, independent of any explanatory exogenous variables; and (c) main effects are linear functions of any explanatory exogenous variables. Under these simplifying assumptions, examination of Eqs. (42) and (43) reveals that the exponents (alternatively half the logarithmic odds ratios) in each case may be expressed as *linear functions* of the exogenous variables and the *scaled* variables u_1, u_2, u_3. Thus, these simplifying assumptions correspond to the assumption of linear structural equations in a simultaneous equations system.[15]

The applicability of the methodology described in this appendix is illustrated in Nerlove and Press, 1974, Section 5. It is shown there, by means of a variety of empirical economic applications, that the above procedures are required in many situations where alternative procedures give misleading, inaccurate, or incorrect results. In particular, in that report we compare the results obtained for dichotomous dependent variables using the linear probability with those obtained using maximum-likelihood methods, and we compare the estimates of the joint probabilities of several dichotomous variables with those obtained by treating all but one of the jointly dependent dichotomous

---

[15] Unlike the latter, however, the qualitative variable case does not, in general, present identification problems under these simplifying assumptions.

variables as if they were exogenous (conditional estimators). Empirical application of this model has not yet been made in the context of measuring technological change.

BIBLIOGRAPHY TO APPENDIX

Ashford, J. R., and R. R. Sowden, "Multivariate Probit Analysis," *Biometrics*, Vol. 26, 1970, pp. 535-546.

Ashton, W. D., *The Logit Transformation*, New York: Hafner Publishing Company, Inc., 1972.

Bartlett, M. S., "Contingency Table Interactions," *Supplement to the Journal of the Royal Statistical Society*, Vol. 2, 1935, pp. 248-252.

Berkson, J., "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, Vol. 39, 1944, pp. 357-365.

-----, "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function," *Journal of the American Statistical Association*, Vol. 48, 1953, pp. 565-599.

-----, "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function," *Journal of the American Statistical Association*, Vol. 50, 1955, pp. 130-161.

Birch, M. W., "Maximum Likelihood in Three-Way Contingency Tables," *Journal of the Royal Statistical Society*, Ser. B, Vol. 25, 1963, pp. 220-233.

Bishop, Y. M. M., "Full Contingency Tables, Logits, and Split Contingency Tables," *Biometrics*, Vol. 25, 1969, pp. 383-400.

Bliss, C. I., "The Method of Probits--A Correction," *Science*, Vol. 79, 1934, pp. 409-410.

Bock, R. D., "Estimating Multinominal Response Relations," *Contributions to Statistics and Probability: Essays in Memory of S. N. Roy*, Chapel Hill, N. C.: University of North Carolina Press, 1970.

Box, M. J., D. Davies, and W. H. Swann, *Non-Linear Optimization Techniques*, Edinburgh: Oliver and Boyd, Ltd., 1969.

Cornfield, J., and N. Mantel, "Some New Aspects of the Application of Maximum Likelihood to the Calculation of the Dosage Response Curve," *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 181-210.

Cox, D. R., *The Analysis of Binary Data*, London: Methuen, 1970.

Fienberg, S. E., "The Analysis of Incomplete Multiway Contingency Tables," *Biometrics*, Vol. 28, 1972, pp. 177-202.

-----, Y. M. M. Bishop, P. W. Holland, and F. Mosteller, *Discrete Multivariate Analysis: Theory and Practice*, forthcoming.

Finney, D. J., *Probit Analysis*, Cambridge, England: Cambridge University Press, 1947.

Fisher, R. A., "The Case of Zero Survivors," Appendix to C. I. Bliss, "The Calculation of the Mortality Curve," *Annals of Applied Biology*, Vol. 22, 1935, pp. 164-167.

Goldberger, A. S., *Econometric Theory*, New York: John Wiley & Sons, Inc., 1964.

Goodman, L. A., "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries," *Journal of the American Statistical Association*, Vol. 63, 1968, pp. 1091-1131.

-----, "On Partitioning χ^2 and Detecting Partial Association in Three-Way Contingency Tables," *Journal of the Royal Statistical Society*, Ser. B., Vol. 31, 1969, pp. 486-498.

-----, "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications," *Journal of the American Statistical Association*, Vol. 65, 1970, pp. 226-256.

-----, "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," *Technometrics*, Vol. 13, 1971a, pp. 33-61.

-----, "Partitioning of Chi-Square, Analysis of Marginal Contingency Tables, and Estimation of Expected Frequencies in Multidimensional Contingency Tables," *Journal of American Statistical Association*, Vol. 66, 1971b, pp. 339-344.

-----, "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," *American Sociological Review*, Vol. 37, 1972a, pp. 28-46.

-----, "A General Model for the Analysis of Surveys," *American Journal of Sociology*, Vol. 77, 1972b, pp. 1035-1086.

-----, "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys," unpublished, August 1972c.

Grizzle, J. E., "Multivariate Logit Analysis," *Biometrics*, Vol. 27, 1971, pp. 1057-1062.

-----, C. F. Starmer, and G. G. Koch, "Analysis of Categorical Data by Linear Models," *Biometrics*, Vol. 25, 1969, pp. 489-504.

Haberman, S. J., *The General Log-Linear Model*, Ph.D. dissertation, Department of Statistics, The University of Chicago, September 1970.

Halperin, M., W. C. Blackwelder, and J. I. Verter, "Estimation of the Multivariate Logistic Risk Function: A Comparison of the Discriminant Function and Maximum Likelihood Approaches," *Journal of Chronic Diseases*, Vol. 24, 1971, pp. 125-158.

Harman, A. J., *Fertility and Economic Behavior in the Philippines*, The Rand Corporation, RM-6385-AID, 1970.

Hodges, J. L., Jr., "Fitting the Logistic by Maximum Likelihood," *Biometrics*, Vol. 14, 1958, pp. 453-461.

Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, Vol. 11, 2d ed., London: Charles Griffin & Co. Ltd., 1961.

McFadden, D., *A Disaggregated Behavior Model of Urban Travel Demand*, Charles River Associates, Inc., Cambridge, Mass., prepared under contract No. FH-11-7566 for the Federal Highway Administration, U.S. Department of Transportation, Final Report, March 1972.

McGillivray, R. G., "Estimating the Linear Probability Function," *Econometrica*, Vol. 38, 1970, pp. 775-776.

Mangahas, M., *An Economic Analysis of the Diffusion of New Rice Varieties in Central Luzon*, Ph.D. dissertation, University of Chicago, 1970.

Mantel, N., "Models for Complex Contingency Tables and Polychotomous Response Curves," *Biometrics*, Vol. 22, 1966, pp. 83-110.

Mosteller, F., "Association and Estimation in Contingency Tables," *Journal of the American Statistical Association*, Vol. 63, 1968, pp. 1-28.

Nerlove, Marc, and S. James Press, *Univariate and Multivariate Log-Linear and Logistic Models*, The Rand Corporation, R-1306-EDA/NIH, December 1973.

Orcutt, G. H., M. Greenberger, J. Korbel, and A. M. Rivlin, *Microanalysis of Socioeconomic Systems: A Simulation Study*, New York: Harper and Brothers, 1961.

Pearson, K., "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such that It Can Be Reasonably Supposed To Have Arisen from Random Sampling," *Philosophy Magazine*, Ser. 5, Vol. 50, 1900, pp. 157-172.

Press, S. J., *Applied Multivariate Analysis*, New York: Holt, Rinehart & Winston, Inc., 1972.

Roy, S. N., and M. A. Kastenbaum, "On the Hypothesis of No Interaction in a Multi-way Contingency Table," *Annuals of Mathematics and Statistics*, Vol. 27, 1956, pp. 749-757.

Schultz, Henry, *The Theory and Measurement of Demand*, Chicago: University of Chicago Press, 1938.

Schultz, T. Paul, and Julie DaVanzo, *Analysis of Demographic Change in East Pakistan: A Study of Retrospective Survey Data*, The Rand Corporation, R-564-AID, 1970.

Smith, V. K., and C. J. Cicchetti, "Estimation of Linear Probability Models with Dichotomous Dependent Variables," Resources for the Future, Inc., 1972 (mimeo.).

Theil, H., "A Multinominal Extension of the Linear Logit Model," *International Economic Review*, Vol. 10, 1969, pp. 251-259.

Theil, H., "On the Estimation of Relationships Involving Qualitative Variables," *American Journal of Sociology*, Vol. 76, 1970, pp. 103-154.

-----, *Principles of Econometrics*, New York: John Wiley & Sons, Inc., 1971, pp. 628-636.

Tobin, J., "The Application of Multivariate Probit Analysis to Economic Survey Data," Cowles Foundation Discussion Paper No. 1, December 1, 1955.

-----, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, Vol. 26, 1958, pp. 24-36.

Truett, J., J. Cornfield, and W. Kannel, "A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham," *Journal of Chronic Diseases*, Vol. 20, 1967, pp. 511-524.

Walker, S. H., and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," *Biometrica*, Vol. 54, 1967, pp. 167-179.

Warner, S. L., "Multivariate Regression of Dummy Variates Under Normality Assumptions," *Journal of the American Statistical Association*, Vol. 58, 1963, pp. 1054-1063.

Zellner, A., and T. H. Lee, "Joint Estimation of Relationships Involving Discrete Random Variables," *Econometrica*, Vol. 33, 1965, pp. 382-394.