ALVIN J. SILK and GLEN L. URBAN*

The substantial failure rate of new packaged goods in test markets has stimulated firms to seek improved methods of pre-test-market evaluation. A set of measurement procedures and models designed to produce estimates of the sales potential of a new packaged good before test marketing is presented. A case application of the system also is discussed.

Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology

INTRODUCTION

Test marketing is a familiar step in the development of new packaged goods—i.e., branded, low-priced, frequently purchased consumer products. Experimental launchings of new products are intended to expose problems that otherwise would be undetected until full-scale introductions are underway. Although test marketing is commonplace, deciding if and when it should be used in particular cases is a perplexing and controversial management problem. The substantial failure rate historically observed among new packaged goods placed in test markets, plus the high and ever-rising direct cost of such activities, have stimulated firms to seek ways to perform more thorough evaluations of new products *before* embarking on test marketing programs.

The purpose of this article is to report progress in development of a measurement and model system, called ASSESSOR, designed to estimate the sales potential of new packaged goods before they are test marketed. The ultimate aim of such a capability is to reduce the incidence of new product failures in test markets and thereby to effect savings in the total cost of new product development.

*Alvin J. Silk and Glen L. Urban are Professors of Management Science, Sloan School of Management, Massachusetts Institute of Technology. The authors are indebted to Calvin Hodock for important discussions of this subject and essential marketing research support. Special acknowledgement is due Robert Klein who contributed many valuable suggestions and much expertise in data collection and analyses to this project. Helpful comments on a draft were given by John Hauser and two *JMR* referees.

First some data bearing on test market failure rates and costs are reviewed and current pre-test-marketing evaluation methods are examined briefly. After the particular objectives of ASSESSOR are set forth, the measurement methodology, model structure, and estimation procedures used are described. The first application of the system then is discussed in some detail. Finally, some results obtained from subsequent studies are reviewed briefly and the limitations of the approach are considered.

PROBLEM DESCRIPTION

Test Marketing

Manufacturers of packaged goods have come to rely on a fairly common set of measurement methods for assessing consumer response to a new product. The typical approach includes (1) concept and positioning tests, (2) product usage tests, and (3) test marketing [34; 47, Chapter 4]. The last step constitutes the final integration and evaluation of the product formulation and the various elements of the marketing plan designed to implement the desired positioning strategy.

The design and scale of test market operations for new products depend on the specification of purpose in terms of estimation and experimentation [42]. The objective of test marketing is sometimes primarily to obtain an estimate of the market share and/or sale volume that would be realized if the new product were launched nationally. In other cases the aim may be to evaluate alternative marketing mix strategies, and hence the test marketing program involves a true experiment. A recent survey of the test marketing practices of "28 major consumer grocery and drug product companies" in the U.S. found that the "norm" was to run a test market in three areas for 10 or 11 months

Journal of Marketing Research
Vol. XV (May 1978), 171-91

[33]. Over a three-year period, these firms had averaged three test marketing programs each per year. The costs of such efforts are considerable and have been mounting. In 1967 the "going rate" for a year-long test in several markets was reported to be $500,000 [61, p. 45]. Today the comparable figure appears to be nearly a million dollars and the authors are familiar with several three-city test marketing programs that involved outlays of $1,500,000.

Even more than the costs, what has motivated closer scrutiny of test marketing practices is recognition of the distressingly high probability that such an undertaking will lead to the detection of a new product failure rather than a success. A review of the limited data available suggests that either outcome is equally likely. In 1961 and again in 1971, the A. C. Nielsen Company reported the "success ratio" of new brands (health and beauty aids, household and grocery products) that had been test marketed through their facilities [50]. The 1961 study included 103 new brands and the 1971 covered 204 items. "Success" was defined by the "manufacturer's judgment of each brand's performance in test"—namely whether or not the brand was launched nationally. Brands withdrawn from test markets or not introduced nationally were considered "failures." By these criteria, only about half of the new brands test marketed in these two periods were successes (54.4% in 1961, 46.6% in 1971). Similarly, the aforementioned survey of the test marketing practices of 28 major consumer grocery and drug product companies found that in 46% of the 54 specific test market experiences covered by the study, test market sales "fell short of management expectations" [33]. In contrast, Buzzell and Nourse [16, p. 100] observed in their study of the food industry that only 32% of 84 "distinctly new food products" developed in the 1954–1964 period were discontinued after test marketing. This somewhat lower failure rate is probably related to the special character of the sample of products studied—i.e., all were "substantially different in form, ingredients, or processing methods from other products previously marketed by a *given* company" [16, p. 96]. At the individual firm level, 10-year test market success rates of 46% and 60% have been reported for General Foods in the U.S. [5, p. 50] and Cadbury in the U.K., respectively [17, p. 98]. Thus, failure rates ranging from 40 to 60% roughly bracket the publicly reported record of test market experience in the packaged goods field.

Besides being an expensive means of detecting new product failures, test marketing involves other problems. First, the test market performance of a new product can be monitored by competitors and provides them with information and time needed to plan a response. Second, the external validity or "projectability" of test market results to subsequent national performance has long been a subject of debate and controversy [1]. For example, A. C. Nielsen Company compared the first year national market share position of 50 new brands with their first year test market performance and concluded that "the odds are about 50–50 that the national performance will match test results within ±10%" [51, p. 4]. This kind of straightforward comparison assumes that test market conditions with respect to such factors as promotional and distribution support and competitive activity were representative of circumstances in the national market. Such an assumption is rarely, if ever, tenable and experience indicates that the predictive accuracy of test-market-based forecasts can be improved markedly by adjusting for discrepancies between test and national conditions with the help of a model that accounts for the dynamics of the new product response process. Competitors have been known to take deliberate retaliatory actions to disrupt another firm's test markets which make it extremely difficult to untangle the results even by complex, model-based analyses [66].

Pre-Test-Market Evaluation

Packaged goods manufacturers have sought in a variety of ways [17] to cope with the high incidence of new product failures in test markets, difficulties in projecting test market results, and the heavy cost of such activities. The most logical place for improvements is the early stages of the new product development process. More effective search and screening procedures can increase the productivity of development and test efforts and diminish the likelihood that a failure will not be detected until the test marketing stage. In recent years new measurement methods and models have been developed and used to facilitate concept generation, refinement, and evaluation [31, 53, 60, 62, 72]. Application of these techniques is intended to lead to better concepts and products, but does not ordinarily obviate test marketing.

Attention also has been directed to making more careful forecasts of expected test market results before the launching of such operations. The sales and market share observed over time for new, frequently purchased consumer products tend to follow a consistent general pattern that can be understood in terms of the level of cumulative trial the new brand achieves and the rate of repeat purchasing it is able to sustain [52]. As the diffusion process proceeds, trial and repeat purchases move toward steady-state levels giving rise to an equilibrium market share and sales rate. Several models have been developed which use early test market or introductory data to forecast equilibrium share and volume [41, Chapter 17]. Certain of these models have been used to arrive at pre-test-market predictions of equilibrium share employing inputs derived from concept and usage tests for the new product plus data for analogous products and/or judgment [68, 69]. Claycamp and Liddy [22] carried this idea a step further and built a regression model to predict trial and repeat purchase levels before the launch of a test market from a set of controllable and uncontrollable variables measured by a mixture of judgmental ratings and consumer test results. The model was estimated and tested with data obtained from 58 new product introductions that covered 32 different types of packaged goods. Eskin and Malec [28] report progress in developing a model which extends the Claycamp and Liddy work in important ways. Some firms have developed similar models using historical data on new product introductions for more narrowly defined product categories [6; 47, p. 94–100]. Though the evidence reported bearing on the forecasting ability of this approach is encouraging [6, 22, 69], the use of such cross-sectional models is always surrounded by uncertainty about the universe of new products and market conditions over which the parameter estimates can be expected to remain stable [26, 64].

Many packaged goods manufacturers have turned to lower cost alternatives to the traditional multi-area test market as a means of reducing expenditures on new product research [e.g., 17]. Several varieties of scaled-down or "controlled" test markets have been used [1]. These operations typically involve fewer and/or smaller areas, but allow more control over some marketing mix variables than do regular test markets. However, the costs remain substantial (expenditures of $100,000 are common) and the projectability of results to the total market is controversial [47, p. 48]. A related but essentially different approach operative in Western Europe, the "mini test market" [29, 57], is discussed briefly in the section on design considerations.

Another pre-test-market method for evaluating new packaged goods is the "laboratory" or "simulated" test market. The basic design concept is to simulate the awareness-trial-repeat purchase process by controlled laboratory and product usage tests. Measurements obtained at several points in this process are used to predict steady-state market share for the new brand and to provide diagnostic information. These ideas form the basis of the work reported here. Brief mention of previous applications of this type of combined laboratory-use test design in commercial marketing research can be found in the literature [47, p. 44, 59; 49, p. 77-9, 183-5; 64], and several firms are known to offer such services [64]. However, the only detailed account of comparable work known to the authors is in an unpublished paper by Burger [15] who describes the COMP system developed in conjunction with Elrick and Lavidge, Inc. The specific measurements, models, and estimation procedures used in the present study are very different from Burger's.

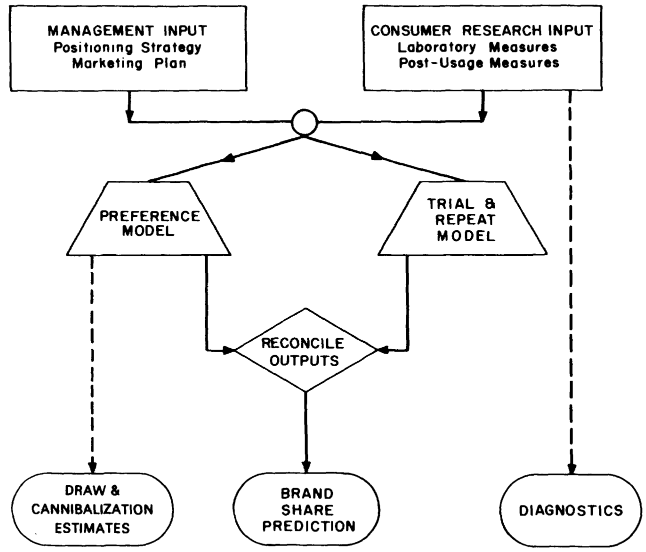
Objectives and Structure of ASSESSOR

ASSESSOR is a set of measurement procedures and models designed to aid management in evaluating new packaged goods before test marketing when a positioning strategy has been developed and executed to the point where the product, packaging, and advertising copy are available and an introductory marketing plan (price, promotion, and advertising) has been formulated. Given these inputs, the system is intended to:

1. Predict the new brand's equilibrium or long-run market share.
2. Estimate the sources of the new brand's share—"cannibalization" of the firm's existing brand(s) and "draw" from competitors' brands.
3. Produce actionable diagnostic information for product improvement and the development of advertising copy and other creative materials.
4. Permit low cost screening of selected elements of alternative marketing plans (advertising copy, price, and package design).

Figure 1 shows the overall structure of the system developed to meet these requirements. The critical task of predicting the brand's market share is approached through two models—one relates preference to purchase probability and the other is a straightforward flow representation of the trial-repeat process. The two models are similar in structure, but are calibrated in different ways. Convergent results should strengthen confidence in the prediction whereas divergent outcomes signal the need for further analyses to identify sources of discrepancies and to provide bases for reconciliation. The measurement inputs required for both models are obtained from a research design involving laboratory and usage tests. The key outputs are a market share prediction plus diagnostic information which can be used to make a decision as to the brand's future. Several outcomes are possible. A poor showing may lead to either termination or further developmental efforts. If the performance is satisfactory, plans for test marketing can proceed. Very favorable results could lead to an immediate launching of the brand, particularly if the capital investment risked in the introduction is small and/or if the threat of competitive entry is imminent.

Figure 1
STRUCTURE OF THE ASSESSOR SYSTEM



RESEARCH DESIGN AND MEASUREMENT

An Overview of the Design

The measurement inputs required to develop the desired diagnostic information and predictions for ASSESSOR are obtained from a research design structured to parallel the basic stages of the process of consumer response to a new product. Table 1 outlines the essential features of the design and identifies the main types of data collected at each step. To simulate the awareness-trial stages of the response process, a laboratory-based experimental procedure is used wherein a sample of consumers are exposed to advertising for the new product and a small set of the principal competing products already established in the market. Next, the consumers enter a simulated shopping facility where they have the opportunity to purchase quantities of the new and/or established products. The ability of the new product to attract repeat purchases is assessed by one or more waves of followup interviews with the same respondents conducted after enough time has passed for them to have used or consumed a significant quantity of the new product at home.

Procedures

The laboratory phase of the research is executed in a facility in the immediate vicinity of a shopping

Table 1
RESEARCH DESIGN AND MEASUREMENT

Design	*Procedure*	*Measurement*
O_1	Respondent screening and recruitment (personal interview)	Criteria for target group identification (e.g., product class usage)
O_2	Premeasurement for established brands (self-administered questionnaire)	Composition of "relevant set" of established brands, attribute weights and ratings, and preferences
X_1	Exposure to advertising for established brands *and* new brand	
$[O_3]$	Measurement of reactions to the advertising materials (self-administered questionnaire)	Optional, e.g., likability and believability ratings of advertising materials
X_2	Simulated shopping trip and exposure to display of new and established brands	
O_4	Purchase opportunity (choice recorded by research personnel)	Brand(s) purchased
X_3	Home use/consumption of new brand	
O_5	Post-usage measurement (telephone interview)	New brand usage rate, satisfaction ratings, and repeat purchase propensity; attribute ratings and preferences for "relevant set" of established brands plus the new brand

O = Measurement.
X = Advertising or product exposure.

center. "Intercept" interviews (O_1) are conducted with shoppers to screen and recruit a sample of consumers having attributes that characterize the target market for the new product. The schedule of this work is staggered over time to reduce the opportunity for obvious kinds of self-selection biases to affect the respondents drawn into the study. Additional control over sample composition can be exercised by carrying out the field work at several different locations chosen to attain the heterogeneity and quotas desired in the final sample. Considerable flexibility is possible here because elaborate facilities and arrangements are not required. Studies done to date typically have involved samples of approximately 300 persons.

Upon arriving at the laboratory facility, respondents are asked to complete a self-administered questionnaire that constitutes the before measurement (O_2). Individually or in pairs, respondents then proceed to a separate area where they are shown a set of advertising materials (X_1) for the new brand plus the leading established brands. Ordinarily, respondents are exposed to five or six commercials, one per brand, and the presentation order is rotated for different groups to avoid any systematic position effects. Measurement of reactions to the advertising materials (O_3) is done next if such information is desired for diagnostic purposes. Dropping this optional feature of the design eliminates a potential source of unwanted reactive effects on respondents' subsequent behavior.

The final stage of the laboratory experiment takes place in a simulated retail store where participants have the opportunity to make a purchase. When first approached, they are told that they will be given a fixed amount of compensation for their time—typically about $2.00 but always more than the sum needed to make a purchase. In the lab they are informed that they may use the money to purchase any brand or combination of brands in the product category they choose, and that they can keep any unexpended cash. They then move to an area where quantities of the full set of competing brands including the new one are displayed and available for inspection (X_2). Each brand is priced at a level equal to the average price at which it is being sold regularly in mass retail outlets in the local market area. The brand (or brands) selected by each participant is (are) recorded by one of the research personnel (O_4) at the checkout counter. Although respondents are free to forego buying anything and to retain the full $2.00, most do make a purchase. For example, the proportion of participants making a purchase observed in two separate studies of deodorants and antacids were 74% and 64%, respectively. Those who do not purchase the new brand are given a quantity of it free after all buying transactions have been completed. Note that this procedure parallels the common practice of effecting trial usage by the distribution of free samples. A record is maintained for each respondent as to whether he or she "purchased" or was given the new brand to allow assessment of whether responses on the post-usage survey are affected differentially by trial purchase versus free sampling.

The post-usage survey (O_5) is administered by telephone after enough time has passed for usage experience to have developed. The specific length of the pre-post measurement interval is determined by the estimated average usage rate for the new product. Respondents are offered an opportunity to make a repurchase of the new brand (to be delivered by mail) and respond to essentially the same set of perception and preference measurements that was used in the before or pre-measurement step (O_2) except that they now rate the new brand as well as established ones. Familiarity with the questionnaire gained from previous exposure makes it feasible to re-administer the instruments in telephone interviews.

Some shrinkage in sample size inevitably occurs between the laboratory session and the post-usage survey. Two important varieties of attrition occur. First, some proportion of persons who participated in the laboratory session will be excluded from the telephone survey as a result of having moved, being away from home, refusing to be interviewed, etc. A second source of sample attrition is respondents who report in the post-usage survey that they have not used the supply of the new product they previously had purchased in the lab store or had been given. In the deodorant study, 16.7% of the original laboratory sample could not be reinterviewed and another 16.7% had not used the product. The general policy has been to continue reinterview efforts until a sample of users of the new product is obtained which includes at least two-thirds of the original set of respondents. Subjects not responding to the post-usage survey can be compared with those who do with respect to information about such factors as last purchase brand share and usage rate obtained from the before measurement (O_2) in order to detect the presence of systematic biases in the post-usage sample due to experimental attrition.

Measurement Instruments

Table 1 identifies the key measures obtained at various points in the design. Certain nonstandard features of the methods require additional discussion. Allaire [3] has shown that measurement of perception and preference structures can be distorted by including unfamiliar stimuli in the set of alternatives judged. Following his methodological recommendation, the authors ask each respondent to provide perception and preference ratings only for those brands that compose his or her "relevant set" of alternatives—i.e., that subset of available brands which are familiar to the respondent regardless of whether they are judged favorably or unfavorably as choice alternatives.[1] Respondents' idiosyncratic relevant sets are revealed by a series of unaided recall questions which identify brands previously purchased or used plus any others considered to be satisfactory or unsatisfactory alternatives.

The size of a typical respondent's relevant set is small in relation to the total number of brands available in the market. Urban's [67] data for seven different categories of packaged goods show that the median relevant set size generally observed is about three brands. Campbell [18] and Rao [59] reported evoked set sizes of approximately the same magnitude for some additional product classes. The smallness of evoked or relevant set sizes is consistent with evidence available as to the number of different brands of packaged goods actually purchased by households. Massy et al. [44, p. 22–4] reported some relevant statistics for a subsample of U.S. households in the J. Walter Thompson panel. During a one-year period, the mean number of different brands purchased per household was 3.3 for regular coffee, 2.6 for tea, and 3.0 for beer. The ranges observed in this quantity for these three product categories were 1–12, 1–8, and 1–11, respectively. Wierenga [70, Chapter 6] investigated some related phenomena using purchase diary data from a panel of 2,000 Dutch households. He found that although a total of 29 different brands accounted for 85% of the total volume of margarine purchased, the mean number of brands purchased per household over a two-year period was only 4.26. The comparable figures for beer and an unidentified food product were eight and 14 brands available, respectively, with 2.57 and 2.88 being the average number of brands purchased per household in these two product categories. Table 2 shows the distribution of relevant set sizes for deodorants observed among a sample of 299 respondents. Here again, the median relevant set size is three brands.

After a respondent's relevant set of brands is identified, attribute importance ratings are obtained. Beliefs/perceptions about the extent to which each brand in a respondent's relevant set offers these attributes also are elicited by means of bipolar satisfaction scales. These two types of data are important components of the diagnostic information provided by the system.

A constant sum, paired comparison procedure is used to assess brand preferences. Several variants of the constant sum approach have been used in marketing research studies and some evidence bearing on the reliability and validity of such measures has been reported. Axelrod [7] employed a constant sum technique as a rating scale device by asking respondents to allocate "11 cards" among a predetermined

Table 2
DISTRIBUTION OF RELEVANT SET SIZES FOR DEODORANTS

Relevant set size	*Percentage of sample (n = 299)*
1 or 2	31.8
3	31.8
4	23.1
5	7.0
6	4.0
7	2.3
	100.0

[1]The term "relevant set" is due to Allaire [3] and is akin to Howard and Sheth's [37] concept of "evoked set." The former consists of familiar alternatives, irrespective of how favorably (or unfavorably) they are evaluated. In contrast, evoked set generally has been interpreted to include only "acceptable" alternatives. For further discussion of this distinction as well as other conceptual refinements and operational definitions different from that used here, see [38, 48].

set of brands so as to indicate the likelihood of their buying each brand. An individual's preference score for a particular brand was simply the number of cards allocated to it. In a complex multistage study, several different awareness and preference measures were compared with respect to their "sensitivity" (ability to detect an effect of advertising exposure in a before-after with control group design), "stability" (aggregate agreement between equivalent samples), and "predictive power" (ability to predict purchases at t_2 from measures obtained at t_1). On the basis of the results obtained, Axelrod recommended use of the constant sum scale to elicit attitude ratings for brands mentioned by consumers in response to an unaided brand awareness question.

Haley [32] reported the results from another comparative study of several attitudinal measures which included a combined paired comparison, constant sum procedure. For all possible pairs of brands, respondents were instructed to divide "10 points" between any two brands so as to reflect their preferences. An individual's preference score was obtained for each brand by summing the points assigned to that brand over all the relevant pairwise comparisons. In relation to the other measures investigated, Haley reported that this method proved superior in its ability to discriminate among brands. Also, it yielded scores whose distribution appeared to be approximately normal.

The findings reported by Axelrod and Haley suggested use of the constant sum technique as a desirable procedure for eliciting preference judgments from consumers. However, in both of these studies and in other marketing research applications, the methods used to estimate scale values for brands from constant sum input data have been of an *ad hoc* variety. In psychophysical measurement where it was first used [65, p. 105–7], constant sum comparative judgments are the basis of an explicit scaling model for which formal estimation methods have been developed. Under the assumption that the subjects can provide *ratio* judgments of paired comparisons between stimuli, Torgerson [65, p. 108–12] devised a least-squares method for estimating ratio scale values. The authors used this form of constant sum, paired comparison scaling to measure a respondent's preferences for his or her relevant set of brands.

The measures of attribute importance weights, brand belief or attribute ratings, and preferences obtained in the before measurement (0_2) are repeated again in the post-usage survey (0_5) but with the new brand added to each respondent's "relevant set" of alternatives. Finally, respondents are given an opportunity to make a mail order repurchase of the new product.

Design Considerations

Selection of the design outlined was influenced by certain operational cost and timing objectives. In particular, the new product management group which initiated this work was seeking a method of producing an evaluation of a new packaged good within a three-month period and at a cost of less than 5% of the typical expenditure required for a test market (i.e., $25,000–50,000). The time and expense required to implement the data collection procedures described here are well within the limits of these design desiderata.

An additional appealing feature of the design is flexibility. It can be expanded for a relatively modest amount of incremental cost to permit evaluations of alternative executions of certain elements of the new product's introductory marketing program—e.g., responses to different commercials can be compared by adding treatment groups to the design, each of which is exposed to a separate commercial.

Mail and home delivery panels are possible alternatives to the approach described. The difficulty of efficiently reaching respondents from the relevant target group and the problem of nonresponse diminish the attractiveness of mail panels. The authors are aware of no published accounts of the use of mail panels for testing new packaged goods. Evidence of the successful use of a home delivery panel in new product testing has been reported by Pymont and his coworkers [21, 29, 57, 58]. Initially developed in the United Kingdom as a "mini test market" facility and subsequently adopted in several other Western European countries, this carefully conceived measurement system involves a continuous panel of households who make purchases from a special door-to-door retail grocery service. Promotional communications and new product introductions are effected by means of controlled print vehicles sent to members of the panel. Charlton et al. [20] analyzed the purchase behavior observed in this environment using Ehrenberg's [25] NBD repeat buying model. They concluded that the brand choice patterns of the mini test panel for established products "are generally like those in real life" in the sense of being consistent with models known to describe purchase behavior under natural conditions. This methodology has been used extensively in Western Europe to evaluate new packaged goods, the obvious attraction being that it offers an efficient means of estimating repeat purchasing for new brands. The latter quantity generally is acknowledged to be the prime determinant of a new brand's success or failure, but the one which is least amenable to rapid and accurate measurement. A high degree of predictive accuracy is claimed for this system and supported by case histories of several applications. Steady-state shares predicted for new brands by the Parfitt-Collins [52] model using estimates of the trial, repeat, and buying rate parameters derived from the mini test panel have been found to be in very close agreement with the comparable share figures observed in concurrent or subsequent normal test markets and/or national introductions [57, 58].

The home delivery panel/mini test market clearly is an appealing alternative to the approach pursued here. It remains an open question whether the former methodology ordinarily would allow the timing and cost criteria established for the present work to be met. The authors have found no documented accounts or other reports of experiences in the U.S. with a mini test market facility. In the United Kingdom, expenditures for new product penetration studies in the mini test market system are said to be "less than 5% of the cost of conventional test marketing" and, on average, about 16 weeks of testing is required [29]. The home delivery arrangement does not permit the television commercial and product display exposure that can be effected in a laboratory facility and hence trial usage can be expected to accumulate more rapidly under the latter approach. However, if extensive usage experience is required for consumers to learn about the new product or if its frequency of purchase differs from that of established brands [8, 27], then one or two waves of post-usage interviews conducted soon after the laboratory session will not provide a reliable basis for estimating its repeat buying rate, and the home delivery panel becomes a preferred and necessary alternative.

Cost and timing considerations aside, the larger design issue is the quality of the measurements. Reference was made heretofore to various steps taken to minimize and/or identify certain threats to validity [19]. One other potential source of confounding effects that merits attention is the use of repeated measures with the same respondents. The available evidence suggests that this is not a troublesome feature of the present design. In a special experimental study undertaken to investigate this issue, the measurement of response to the advertising materials (0_3) was found to have no apparent reactive effect on respondent brand choice behavior observed (0_4) in the simulated shopping trip. Ginter [30] and Winter [73] also investigated this general issue in their laboratory study which involved four consecutive weekly sets of measurements taken before and after exposure to advertising stimuli. They found some indications that the repeated measurements were reactive, but report that these effects were not sufficiently strong or systematic to be problematical [30, p. 33; 73, p. 32].

MODEL STRUCTURE

As shown in Figure 1, two different models are used to generate separate predictions of market share for a new brand. The first relates strength of post-trial preference for the new brand to the probability of purchasing it. The second is a more direct representation of the trial-repeat purchase process. In this section the details of each model are set forth and their structural correspondence is examined. A discussion of how the output of the system is used for strategic management issues follows.

Preference Model

The fundamental problem addressed here is that of predicting market share, an aggregate measure of purchase behavior. The available empirical evidence leads one to favor selection of preference over other attitudinal or behavioral disposition constructs as a simple predictor of brand choice. As was mentioned, Axelrod [7] found the predictive power of preference ratings (obtained by a constant sum procedure) superior to that of a variety of other interview/questionnaire-based evaluative measures for established brands of packaged goods.

In the first of a series of important studies, Pessemier et al. [55] demonstrated that their interval-scaled "dollarmetric" measure of brand preferences [54] obtained in a laboratory setting could be used to develop fairly accurate predictions of the relative frequency of individual consumers' subsequent purchases of established brands under natural conditions over a seven-month period. More recently, Ginter [30] conducted an experimental study of response to a new brand that involved a sequence of four weekly laboratory sessions wherein housewives were exposed to commercials for new brands in two different packaged goods categories and were given the opportunity to purchase them in a simulated shopping trip. Among other things, he found that preference (measured by the same method as that used previously by Pessemier et al.) was a better predictor of purchase of the new brands than a multiattribute attitude model. The several unresolved issues [71] that surround the latter class of models further discourage their use for the present purposes.

On the basis of subsequent work, Bass et al. [11] argue that although preference measures do exhibit significant predictive power, a high degree of accuracy cannot be realized because of measurement error, omitted variables, random exogenous events, etc., and perhaps consumers' "desire for variety." This leads them to the view that "since choice behavior is not constant even when attitudes are unchanging, attitude-based predictions of choice must be probabilistic" [11, p. 541].[2]

A similar orientation has been adopted here: the authors first estimate individual consumers' probabilities of purchasing the new brand from their expressed brand preferences after a period of initial usage of it and then aggregate these probabilities across individuals to obtain an estimate of expected aggregate or total market share.

Luce's probabilistic theory of choice [43] provides a valuable foundation for formulating a model to link

[2]Bass has gone further in developing this position. See [9].

brand preferences to purchase probabilities. Luce has shown that a simple but powerful axiom about choice probabilities implies the existence of a ratio scale for the alternatives. More specifically, the Luce model, written in terms of brand choice probabilities and preferences asserts that:

$$(1) \qquad P_i(j) = \frac{V_i(j)}{\sum_{k=1}^{m_i} V_i(k)}, \qquad V_i(k) > 0,$$

where:

$P_i(j)$ = probability that consumer i chooses brand j,
$V_i(j)$ = consumer i's ratio scaled preference for brand j,
k = 1, ..., j, ..., m_i,
m_i = number of brands in a respondent's relevant set of alternatives.

In the present context, the authors postulate that the observed measures of preference, obtained by the constant sum, paired comparison procedure referred to previously, are related to brand choice probabilities by:

$$(2) \qquad P_i(j) = \frac{[\hat{V}_i(j)]^\beta}{\sum_{k=1}^{m_i} [\hat{V}_i(k)]^\beta}, \qquad \hat{V}_i(j) > 0,$$

where:

$\hat{V}_i(j)$ = estimated preference of consumer i for brand j,
β = parameter to be estimated.

This form of preference model has been used in consumer research by Pessemier et al. [55]. They found that straightforward application of their interval scale preference measure to equation 1 resulted in overprediction of the relative frequency of purchase of less preferred brands. Better fits were realized with equation 2 where a heuristic method was used to obtain an estimate of β that was product class specific, but which applied across all brands and all consumers. Pessemier et al. [55] discussed the application of the exponent β to the preference scores as a means of accounting for noise and discrepancies between laboratory and market conditions. Similarly here, the ratio scaling of preferences to which the constant sum, paired comparison procedure aspires may not be attained and the exact properties of the preference scale rendered by use of the method cannot be directly ascertained. Pessemier and Wilkie [56] pointed out that the transformation implied in equation 2 that equates it to equation 1 is similar to Steven's Power Law [63] used in psychophysical research to relate subjective magnitude to physical magnitude.

Formulation 2 also may be related to McFadden's [46] "random utility model" which he derived as a theory of population choice behavior, building upon Luce's individual choice model. McFadden assumes the utility (μ_{ij}) each member (i) of a utility-maximizing population of consumers has for a choice alternative (j) consists of a measurable (c_{ij}) component and a stochastic element (ε_{ij}), i.e.:

$$(3) \qquad \mu_i(j) = c_i(j) + \varepsilon_i(j).$$

The nonstochastic component is taken to be a function of a vector of attributes describing the alternatives faced by the individual. Assuming the $\varepsilon_i(j)$ are independent Weibull distributed, McFadden shows that Luce's model of individual behavior leads to an econometric specification of the choice probabilities as a multinomial logit model similar to equation 2, i.e.,

$$P_i(j) = \frac{\exp\,[c_i(j)]}{\sum_k \exp\,[c_i(k)]}.$$

Empirical experience also has led to the use of equation 2 in this work. The authors estimate β using the preference scale values for the *established brands* derived from data obtained in the pre-exposure questionnaire (O_2 in Table 1) and information about the last brand which respondents report having purchased. Statistical methods for estimating β are discussed in the next section. Assuming β to be a stable parameter whose value will remain unchanged after introduction of the new brand, and given measures of consumers' preferences for the new brand plus the established brands obtained *after* a period of trial usage of the new brand, it follows from equation 2 that one can predict each individual's probability of purchasing the new brand using:

$$(4) \qquad L_i(t) = \frac{[A_i(t)]^\beta}{[A_i(t)]^\beta + \sum_{k=1}^{m_i} [A_i(k)]^\beta},$$

where:

$L_i(t)$ = probability that consumer i chooses the brand t after having tried the new brand,
t = index for the new brand,
k = index for established brands,
$A_i(t)$ = estimated preference of consumer i for the new brand t after having tried the new brand,
$A_i(k)$ = estimated preference of consumer i for established brand k after having tried the new brand.

Now the predicted probabilities are conditional upon the brand being an element of each consumers' relevant set. To calculate an expected market share for the new brand one must take into account that the new brand will not necessarily become an element of the relevant set of brands for all consumers when it does become available in the market. Therefore,

$$(5) \qquad M(t) = E(t)\,\frac{\sum_{i=1}^{N} L_i(t)}{N},$$

where:

$M(t)$ = expected market share for the new brand t,
$E(t)$ = proportion of consumers who include brand t in their relevant set of alternatives,
$L_i(t)$ = predicted probability of purchase brand t by consumer i, $i = l, ..., N$.

To use equation 5 to forecast the new brand's market share, one first must predict the proportion of consumers, $E(t)$, who will consider the new brand as a relevant alternative. A procedure for estimating this quantity is discussed in the next section.

Where there is substantial variation in consumption among individual consumers, the $L_i(t)$ in equation 5 must be weighted by a usage rate index.

The task of predicting how the new brand will affect the shares of existing brands requires that one obtain their expected market share when equilibrium is reestablished after the launching of the new brand. To do so, one must again recognize that under the new steady-state conditions the market will consist of two subpopulations, distinguishable by the presence or absence of the new brand in their relevant sets. The sizes of these two groups in relation to the total target market will be $E(t)$ and $1\text{-}E(t)$, respectively. The addition of the new brand to a respondent's relevant sets is effected experimentally by the procedures noted previously and so the impact of its inclusion will be manifested in the preferences for the established brands expressed by respondents in the post-usage survey (*0*) *after* exposure to the new brands—i.e., in the quantities $A_i(k)$. In contrast, it seems reasonable to suppose that consumers whose relevant set does not include the new brand will continue to purchase established brands after the new brand is available in the same manner as they did before its entry—i.e., according to the established brand preferences held *before* exposure to the new brand, $\hat{V}_i(k)$, as defined in equation 2. One also assumes that (1) the probability of the new brand being included in a consumer's relevant set is independent of relevant set size and composition or the structure of preferences for established brands, and (2) inclusion of the new brand in a consumer's relevant set does not affect the number or identity of established brands it contains. Using these ideas one derives expected market shares for established brands in the following manner. As in equation 4, if the new brand is present in a consumer's set, the purchase probability for any established brand j will be given by:

$$(6) \qquad L_i(j) = \frac{[A_i(j)]^{\beta}}{[A_i(t)]^{\beta} + \sum_{k=1}^{m_i} [A_i(k)]^{\beta}},$$

$$k = 1, ..., j, ..., m_i,$$

and its share in the submarket of consumers whose relevant set includes the new brand is:

$$(7) \qquad M(j)' = \frac{\sum_{r(j)} L_i(j)}{N},$$

where the summation $\Sigma_{r(j)} L_i(j)$ is over the $r(j)$ individuals who include the established brand j in their relevant sets.

For consumers who do not include the new brand in their relevant sets, the probability of purchasing any established brand j can be obtained from equation 2 and within the subpopulation of all such consumers its market share will be:

$$(8) \qquad M(j)'' = \frac{\sum_{r(j)} P_i(j)}{N}.$$

To obtain the established brand's expected market share in a total market, one weights the "unadjusted" shares (equations 7 and 8) by the relative sizes of the two subpopulations, or:

$$(9) \qquad M(j) = E(t)\, M'(j) + (1 - E(t))\, M''(j),$$

where:

$M(j)$ = expected market share for the established brand j after introduction of the new brand t.

Note that because it follows from equations 4 and 6 that:

$$(10) \qquad \sum_{k=1}^{m_i} L_i(k) + L_i(t) = 1, \qquad i = 1, ..., N,$$

and similarly because from equation 2:

$$(11) \qquad \sum_{k=1}^{m_i} P_i(k) = 1, \qquad i = 1, ..., N,$$

then the expected market shares given by equations 5 and 8 will be logically consistent:

$$(12) \qquad \sum_{k=1}^{m^*} M(k) + M(t) = 1,$$

where:

m^* = total number of existing or established brands.

Comparing the expected market share given by equation 9 for any established brand with its prior share permits one to estimate the impact of the new brand on the existing structure of market shares for established brands.

The key assumptions underlying the model serve to emphasize the conditions under which it can be expected to apply. First, crucial to the Luce-McFadden choice models is the notion of "independence of irrelevant alternatives." Formally, the requirement is that the "ratio of the probability of choosing one alternative to the probability of choosing the other should not depend upon the total set of alternatives available" [43, p. 9]. As discussed elsewhere [23,

p. 150–1; 46, p. 113], this assumption will not hold when the set of alternatives is sufficiently heterogeneous that choices are made in a hierarchical manner, as when a consumer first selects among several product types and then chooses a brand within a particular subcategory. The practice followed here of identifying idiosyncratic relevant sets of alternatives appears to offer some protection against mixing together alternatives that differ markedly in their perceived substitutability. Though it may also be possible to model the structure of a hierarchical choice process separately, attention ultimately must be focused on relatively homogeneous sets of alternatives. Some evidence bearing on the independence-of-irrelevant-alternatives assumption will be discussed in connection with estimation of the preference model.

A second important assumption is the treatment of brand choice as a heterogeneous, stationary, zero-order Bernoulli process [45, Chapter 3]. A survey of the issues and pertinent evidence is given by Bass [10] who emphasizes that stochastic choice models built on these premises are consistent with stable market shares accompanied by substantial brand switching, conditions which are frequently observed together in packaged goods markets. Bass et al. [12] discuss the relationship between heterogeneous, zero-order brand switching models and penetration models like those of Ehrenberg and his coworkers [25]. In addition, they show that, under certain assumptions about how brand preferences are distributed in the population, the Luce choice model leads to a flexible and tractable distribution of purchase probabilities.

Rather than model and measure the dynamics of the adoption process directly, the authors seek to compare equilibrium or steady-state market shares before and after introduction of a new brand, while allowing for heterogeneity in the population of consumers. For the approximation of stationarity to be plausible, market shares for established brands should be constant before the new brand's launch and preferences must have stabilized when the post-usage measures are taken. The latter condition can be checked by repeating the post-usage survey after consumers have acquired additional usage experience with the new product.

Trial-Repeat Model

The steady-state market share a new brand finally achieves can be represented directly as the product of the long-run levels of trial and repeat purchasing it attains. Following Parfitt and Collins [52], one can express market share for the new brand, $M(t)$, by:

$$M(t) = T\,S, \tag{13}$$

where:

T = ultimate cumulative trial rate for the new brand, t (proportion of all buyers in the target group who ever try the new brand),

S = ultimate repeat purchase rate for the new brand, t (new brand's share of subsequent purchases in the product category made by buyers who have ever made a trial purchase of the new brand).

This model has been used extensively [2, 52] to forecast equilibrium shares ($M(t)$) for new brands using extrapolations of early test market measurements to estimate the ultimate trial (T) and repeat purchase (S) rates of equation 13. Here the authors employ a model previously used by Urban [67] which decomposes these two quantities slightly. The purpose is to represent the influence of certain marketing policy variables on consumer response in a simple fashion and at the same time make use of measurements obtained from the laboratory and post-usage studies.

One assumes that trial comes about in one of two ways: (1) receipt and use of free samples or (2) initial purchases. The incidence of first purchase of the new brand is taken to be dependent on the level of awareness induced by advertising or other forms of promotion and the extent of its retail availability. As an approximation, the probability of becoming aware of the new brand and that of having it available are presumed to be independent. One also assumes that the probability a consumer makes a first purchase is independent of the probability of receipt and use of a sample. Putting these assumptions together, one can model trial by:

$$T = \mathrm{FKD} + CU - (FKD)(CU), \tag{14}$$

where:

F = long-run probability of a consumer making a first purchase of the new brand given awareness and availability of it (i.e., proportion of consumers making a trial purchase in the long run given that all consumers were aware of it and distribution was complete),

D = long-run probability that the new brand is available to a consumer (e.g., proportion of retail outlets that will ultimately carry the new brand weighted by their sales volume in the product category),

K = long-run probability that a consumer becomes aware of the new brand,

C = probability that a consumer will receive a sample of the new brand,

U = probability that a consumer who receives a sample of the new brand will use it.

The various probabilities are averages for the particular target group under consideration. As an estimator of F, one uses the proportion of respondents who purchased the new brand (0_4 in Table 1) in the laboratory on their simulated shopping trip. The next three parameters, K, D, and C, depend on the type and magnitude of marketing effort management plans to use if the brand is test-marketed or otherwise launched. Thus, a prime determinant of the level of awareness (K) for the new brand is the amount to be spent for media advertising, whereas the extent of availability (D) depends on how much salesforce and promotional activity will be directed at the retail trade. The translation of the introductory marketing plan into estimates of K and D is accomplished by informal means, drawing upon managerial judgment as well as results and experience obtained with similar products. Analyses of certain types of historical data also can be helpful as, for example, in formulating a relationship between brand awareness and media expenditures or coverage. Estimation of the sampling coverage parameter (C) is straightforward, given knowledge of the scale of sampling program planned. Previous research with similar products or a small experiment can be used to estimate sample usage (U).

Urban [67] models the other quantity in equation 13, S, as the equilibrium share of a first-order, two-state Markov process:

$$S = \frac{R(k,t)}{1 + R(k,t) - R(t,t)}, \tag{15}$$

where the transition probabilities are defined as follows.

$R(k,t)$ = probability that a consumer who last purchased *any* of the established brands (k) will switch to the new brand (t) on the next buying occasion,

$R(t,t)$ = probability that a consumer who last purchased the new brand will repurchase it on the next buying occasion.

Estimates of $R(k,t)$ and $R(t,t)$ are derived from measurements obtained in the post-usage survey (0_5 in Table 1). The proportion of respondents who make a mail order repurchase of the new brand when given the opportunity to do so is taken as an estimate of $R(t,t)$. To estimate $R(k,t)$ for those who do *not* repurchase the new brand in this situation one uses their preference measurements for the new and relevant established brands obtained from them in the post-usage survey. Probabilities of purchasing the new brand are computed for each such individual by equation 4 and their average value is taken as an estimator of $R(k,t)$.

It is sometimes observed empirically that respondents who "purchased" the new brand in the laboratory experiment differ from those who received it as a free sample with respect to their repeat rates, S. In this case, separate repeat rates are calculated and applied to the appropriate trial components in equation 14 to adjust for the difference.

Applying the inputs to equations 14 and 15 gives estimates of the ultimate trial (T) and repeat (S) rates, respectively, which are then simply multiplied together as indicated by equation 13 to calculate the expected long-run market share for the new brand.

This trial-repeat model is clearly a very simplified representation of the new product response process. Some tests of the adequacy of the model's overall structure have been reported by Urban [67]. He derived the various inputs required by the trial and repeat equations for several new products from studies of their test markets or national introductions. He then compared the ultimate trial and repeat rates and equilibrium market shares predicted by the model with the values of these quantities that actually had been observed. For each of the half dozen cases examined, he found the observed and predicted values to be in very close agreement.

In terms of its complexity, the foregoing model has proved to be adequate for the level of detail typically specified in introductory marketing plans at the stage of a new brand's development where the decision to test market or not is under consideration. An important assumption implicit in the model is that the frequency of purchase of the new brand will be the same as that for established brands. This assumption can be relaxed somewhat by weighting the ultimate repeat rate (S) by an index that reflects the new brand's usage rate in relation to that for established brands [52,67]. Clearly the latter is at best a crude adjustment and situations can arise where, if the required measures can be obtained, it will be desirable to use one of the available models [e.g. 66] that allows the adoption process to be represented in greater detail.

Structural and Output Comparisons

The expression for market share developed from the individual preference-purchase probability model (equation 5) is structurally equivalent to that defined in terms of trial and repeat purchase levels (equation 13). In the former case, market share is the product of the relevant set proportion ($E(t)$) and the average conditional probability of purchasing the new brand ($\Sigma_{i=1}^{N} L_i(t)/N$). In the latter case, market share is the product of the cumulative trial proportion (T) and the share which repeat purchases of the new brand represent of subsequent buying by previous triers (S).

Though not precisely identical, "relevant set" and "trial" are operationally very similar constructs in the present context. As noted in the discussion of measurement procedures, the composition of a consumer's relevant set is determined by responses to a series of questions about which brands he/she has ever *used* or would consider using or not using. Thus, one would expect to find that brands so evoked for the most part are accounted for by past usage or "trial," and empirically this is the case. For example, in separate studies of three different product classes, 90% or more of all brands respondents deemed relevant were identified on the basis of usage-related questions.

The quantities $\Sigma\ L_i(t)/N$ and S are both average conditional probabilities or shares of repeat purchases. However, they are distinguished conceptually in that the former are obtained from a zero-order individual level model whereas the latter arise from an aggregate first-order Markov process. Despite these differences, it is often difficult to distinguish between these two types of models, each of which may yield satisfactory results [45]. For example, aggregation over heterogeneous consumers will tend to *overestimate* the true order of the process [45, Chapters 3 and 4]. However, Kesavan and Srinivasan [40] have shown that aggregation of several brands into a single "other" brand category (as is done here) will tend to *underestimate* the true order of the process and can lead to biased steady-state market share predictions. A second difference is that the average purchase probability obtained from the preference model ordinarily will reflect some effects of in-store promotion whereas these are not incorporated explicitly in the estimate of the repeat rate. The parameter of the preference-purchase probability model is estimated from data pertaining to the purchase of established brands which are supported by some level of in-store promotional activity, but no provision for such an effect is made in the repeat submodel (equation 15).

The submodels and measures used to arrive at estimates of these conceptually similar quantities are distinct. Whereas the trial and repeat proportions are based on essentially *direct* observations of these quantities obtained under controlled conditions, the relevant set proportion and the average conditional purchase probability are estimated *indirectly* from other measures. Coming from the same research design, the measurement inputs for both models are affected by common sources of methods variance. Nonetheless, because of differences in the submodels and their respective inputs, agreement between the two market share predictions is not a builtin or guaranteed feature of these approaches and hence it is possible to make a meaningful check for convergence here.

Finding that the two models do yield outputs that are in close agreement can serve to strengthen confidence in the prediction. In contrast, divergent forecasts trigger a search for and evaluation of possible sources of error or bias that might account for the discrepancy. The first step is to compare the relevant set proportion ($E(t)$) and trial (T) estimates. Lack of agreement here could imply that the assumptions concerning awareness (K) and retail availability (D) are not compatible with those made implicitly or explicitly in estimating the relevant set proportion ($E(t)$) as, for example, when the latter is based on a regression of relevant set proportions on awareness levels for established brands. Given that these assumptions did appear compatible, then the possibility of measurement bias in the conditional trial probability (T) would be investigated.

After reconciliation of the trial and relevant set estimates, attention is focused on the values of the conditional purchase probability and the repeat rate. In comparing these quantities, it is important to keep in mind that effects of in-store promotional support are not represented in the repeat rate estimate derived from the post-usage interview. For product classes having substantial in-store promotional programs, upward adjustments in these initial estimates of repeat rates are necessary and justifiable. In the end, some judgment may have to be exercised to reconcile differences that arise, but that process is facilitated by careful consideration of the structural comparability of the two models.

Predictions and Marketing Plans

Prediction of a new brand's market share must reflect plans for the marketing program to be used in the future test market or launch. Frequently at this pre-test market stage management is interested in evaluating some variations in the introductory marketing mix for the new brand. The trial-repeat model can be used to advantage in performing some rough and ready simulations of the effects of certain kinds of marketing mix modifications. Some of the changes or alternatives management may wish to consider can be approximated by judgmentally altering parameter levels. For example, increasing the level of advertising spending could be represented by raising the awareness probability, K, in equation 14. Differences in sampling programs could be handled similarly by modifying the C and U parameters. Other types of changes, as in advertising copy or price, that affect the conditional first purchase probability, F, can be measured by expanding the research design shown in Table 1 to allow observation of the differential effects on trial purchases made in the controlled shopping environment for alternative price or copy treatments.

After examination of the impact of strategic changes, profitability measures can be calculated for the market share estimates. On the basis of these inputs, management must decide whether or not to test market the new brand.

ESTIMATION

At several points in the preceding discussion, reference is made to how data obtained from the laboratory and post-usage phases of the consumer research could be related to the models' parameters and input requirements. For the most part, this is a straightforward task involving only simple computations. However, estimation of the preference scale values, the parameter of the purchase probability model, and the relevant set proportion is somewhat more complex and is discussed in detail hereafter.

Preference Scaling

Data obtained by the constant sum, paired comparison method are used to estimate a vector of brand preference scale values for each respondent by the least-squares procedure proposed by Torgerson [65, p. 109–12]. Respondents' preferences are scaled twice, before and after using the new brand. The "before" scaling is carried out with reference to respondents' idiosyncratic relevant sets identified by the premeasurement (0_2 in Table 1) and the "after" scaling (0_5) encompasses the previously determined relevant set of established brands plus the new brand.

Under the assumption that the comparative judgments reported by a subject for stimuli reflect the ratios of their corresponding subjective magnitudes, then the least squares estimate of the stimulus scale has ratio scale properties. That the computed estimates actually attain this level of measurement cannot be verified from the input data and no statistical test for goodness of fit is available. Two types of internal consistency checks which bear on the quality of the preference scale estimates have been performed with data for deodorant and antacid categories obtained from separate samples. First, very few instances of intransitivities in preferences were uncovered when Kendall's method of circular triads was applied to each subject's paired comparison judgments [39, Chapter 11]. The absence of inconsistencies is not a very demanding requirement here inasmuch as transitivity is only a necessary condition for the existence of an ordinal scale [23, p. 13–24] and with typical relevant set sizes of three to five brands, the number of paired comparison judgments required of subjects is most often small. As Torgerson [65, p. 116] suggests, a goodness-of-fit check also was made. The matrix of ratios representing a respondent's original paired comparison judgments was compared with the equivalent matrix calculated from the estimated brand preference scale values for that respondent. For the great majority of respondents, the estimated scale values for an individual's relevant set of m brands could very accurately reproduce the $m(m-1)/2$ observed ratios that person had provided in performing the paired comparison judgments [5].

Estimation of the Purchase Probability Function Parameter

In the pre-exposure interview, the brands last purchased by respondents are identified and preference measures are obtained for their sets of relevant alternatives. This information on last brand purchase and the brand preference scale values are used to estimate the parameter β of the purchase probability model defined in equation 2. Recall from the discussion of the preference model that β is to be estimated across different (established) brands and across respondents. Also, the observations are (dichotomous) purchase events, not probabilities. Now because

$$[\hat{V}_i(j)]^{\beta} = \exp\,[\beta \ln \hat{V}_i(j)],$$

one can write the purchase probability model (equation 2) as:

$$(16)\qquad P_i(j) = \frac{\exp\,[\beta \ln \hat{V}_i(j)]}{\sum_{k=1}^{m_i} \exp\,[\beta \ln \hat{V}_i(k)]}.$$

The form of the expression is that of the multinomial logit model which McFadden [46] derived as a theory of population choice behavior. Maximum likelihood estimation procedures have been developed for this model and McFadden notes that the estimators obtained are asymptotically efficient and normally distributed under "very general conditions" [46, p. 119].[3] This method has been applied widely in economic studies of choice behavior [24] and is used here to estimate the β parameter in equation 16. More specifically, a program developed by Manski and Ben Akiva [13] is employed which uses the Newton-Raphson iterative technique to determine the value of the parameter β which maximizes the following likelihood function:

$$(17)\qquad L = \prod_{i=1}^{N} \prod_{k=1}^{m_i} [P_i(k)]^{\delta_{ik}},$$

where:

$$\delta_{ik} \begin{cases} = 1 \text{ if individual } i \text{ last purchased brand } k. \\ = 0 \text{ otherwise.} \end{cases}$$

Though standard errors and associated t statistics for the β parameters in equation 16 can be obtained, the usual goodness-of-fit measure, the coefficient of determination (R^2), cannot be applied here because the estimated equation predicts probabilities whereas the observed values are purchase events (0, 1 measures). However, Hauser [35, Chapters 10, 36] recently developed useful measures for assessing the fit of this model based on information theory concepts. Hauser views the model (16) as an information system—i.e., the probabilities obtained from the preference model provide information about the choice outcomes. Now the *prior* entropy measures the total uncertainty in the system *before* observation of the preference data. To compute the prior entropy, Hauser proposes that a naïve model be assumed whereby every member of the sample is assigned a probability of purchasing any brand ($P'(k)$) equal to its aggregate market share among the total sample's reported last purchases. Under this assumption, he demonstrates that the prior entropy is given by:

$$(18)\qquad Z = -\sum_{k=1}^{m^*} P'(k) \log P'(k),$$

where:

Z = total uncertainty in the system with m^* alternative brands,
$P'(k)$ = prior probability of choice of brand k, $k = 1, \ldots, m^*$.

After application of the observed data to the preference model (16), the uncertainty is reduced to the *posterior* entropy. Hauser shows that the amount by which the preference data reduce the prior entropy is the *expected information, EI,* provided by the model which is:

$$(19)\qquad EI = \sum_{i=1}^{N} \sum_{k=1}^{m_i} P_i(k) \log \frac{P_i(k)}{P'_i(k)},$$

where the $P_i(k)$ are obtained from equation 16.

[3]The small sample properties of the maximum likelihood estimator of the multinomial logit model are, in general, unknown. However, on the basis of examples and Monté Carlo studies McFadden suggests that the approximation is "reasonably good." See the discussion in [46, p. 119 ff.].

Noting that the prior entropy also can be taken as a measure of how well a perfect model would perform, Hauser proposes that the "usefulness" of the model (16) be assessed by comparing the expected information, *EI*, with the prior entropy, *Z*. Thus an index of the model's usefulness can be defined as the proportion of total uncertainty removed or "explained" by the model:

$$G = \frac{EI}{Z}. \tag{20}$$

Finally, Hauser shows the observed or empirical information, *OI*, is:

$$OI = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{m_i} \delta_{ik} \log \frac{P_i(k)}{P_i'(k)}, \tag{21}$$

where:

$$\delta_{ik} \begin{cases} = 1 \text{ when respondent } i\text{'s last purchase was brand } k, \\ = 0 \text{ otherwise.} \end{cases}$$

He argues that with a large sample, the observed information should be close to its expected value and thus the "accuracy" of the model can be assessed by comparing *OI* and *EI*.

Table 3 shows the results obtained when the maximum likelihood procedure was applied to the preference and last brand purchased data from the deodorant study. Note that the estimated value of β is nearly 10 times its estimated standard error and the model accounts for slightly more than three-quarters of the total uncertainty present as measured by the index *G*. As expected, the value of *OI* is very close to that of *EI*.

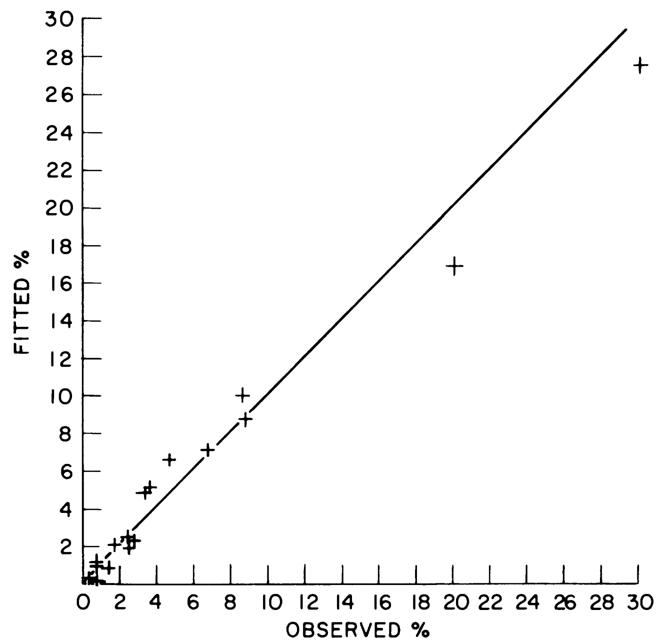
As another check on the adequacy of the fit obtained, the estimated value of β in equation 16 was used to calculate each individual's probability of having last purchased each brand in his/her relevant set. These probabilities were aggregated to calculate the fitted value of each brand's expected share of last purchases. The latter can be compared with the observed shares. Across all 18 brands the mean absolute deviation was found to be .8 of one market share point (percentage). This value can be compared with an average absolute deviation of 2.5 market share points obtained for a "naïve" model whereby an individual has the same probability of purchasing any brand in his/her evoked set—i.e., $P_i(j) = 1/m_i$. Figure 2 shows a plot of the observed and fitted shares. The largest deviations were for the two major brands where the model overpredicted their shares by 2.0 and 3.1 share points, respectively.

In the present context, the "independence of irrelevant

Table 3
MAXIMUM LIKELIHOOD ESTIMATION RESULTS

	Sample size	$\hat{\beta}$	*Standard error*	*Fit indices* *G*	*EI*	*OI*
Total sample	279	2.09	.20	.77	1.72	1.64
By relevant set size						
Two brands	85	1.84	.41	.83	1.56	1.53
Three brands	90	2.75	.49	.84	1.74	1.61
Four brands	65	2.20	.37	.72	1.66	1.56
Five or more brands	39	1.80	.36	.55	1.23	1.12

Figure 2
PLOT OF OBSERVED VERSUS FITTED MARKET SHARES
(N = 18 brands)



alternatives" assumption underlying the Luce-McFadden models implies that β should not vary with relevant set size. To investigate this matter, the model (16) was estimated separately within groups defined by relevant set size. Table 3 shows the results. Some variation in the estimated β can be seen. However, none of the four β estimates is significantly different from the overall or total sample value at the .05 level. Making all possible pairwise comparisons among the four values for the different relevant set sizes, one finds only two of the six differences to be significantly different at the .05 level. The quality of the fit as measured by the *G* index diminishes as the relevant set size increases but the sample sizes for the two largest relevant set size groups are also smaller.

As another test, equation 16 was estimated separately for each pair of brands within the subsample of respondents whose relevant set size was three. None of the β estimates so obtained differed significantly at the .10 level from the value obtained by estimating the parameter across all brands. These results do not appear to indicate any systematic contradiction of the assumption of independence of irrelevant alternatives for these data.

Estimation of the Relevant Set Proportion

Recall that from the preference model one obtains an estimate of the probability of purchasing the new brand that is conditional on its being a relevant choice alternative. Thus, one requires a method of predicting what proportion of consumers in the target group will eventually include the new brand in their relevant sets ($E(t)$ in equation 5).

In the discussion of the comparability of the trial-repeat and preference models, it was noted that for the operational definition used here almost all the brands composing consumers' relevant sets are those with which they report having had some usage experience. Also there tends to be a strong and stable concurrent relationship across brands between aggregate levels of brand awareness and usage [e.g., 14]. Thus similar relationships between relevant set and awareness proportions are suggested, and have been found in the present work. To illustrate, cross-sectional regressions of relevant set proportions ($E(j)$) on unaided brand awareness ($B(j)$) and advertising awareness ($AA(j)$) levels were performed for 18 established brands of deodorants using measures of these variables obtained in the premeasurement questionnaire (O_2 in Table 1). Given that the observations were proportions which varied considerably in magnitude, an arcsin transformation was applied to them as a means of stabilizing the error variance and thereby obtaining efficient estimates from ordinary least squares regressions. The following results were obtained:[4]

$$\text{(22)} \quad \text{Arcsin}\, E(j) = -.599 + \underset{(23.4)}{.901}\, \text{Arcsin}\, B(j) + e(j),$$

$$R^2 = .972, \text{SEE} = 2.39.$$

$$\text{(23)} \quad \text{Arcsin}\, E(j) = 3.91 + \underset{(11.6)}{1.066}\, \text{Arcsin}\, AA(j) + e(j),$$

$$R^2 = .894, \text{SEE} = 4.61.$$

As expected, both brand and advertising awareness appear to covary with the relevant set measure. However, the values of the coefficient of determination (R^2) and the standard error of estimate (SEE) indicate that the brand awareness regression provided a better fit of the data than did the estimated advertising awareness equation. Transforming the estimated values of the arcsin of $E(j)$ from the regression back to proportions and comparing them to their corresponding observed values, one finds the average residual for the brand awareness regression to be .021 and that for the advertising awareness regression to be .041.

To estimate the expected relevant set proportion for the new brand ($E(t)$), one simply applies the level of unaided brand awareness ($B(t)$) which the introductory marketing program is expected to achieve to the brand awareness equation. As noted in the discussion of the trial-repeat model, the level of brand awareness predicted for the new product is largely a judgmental estimate because it depends on the nature and magnitude of marketing effort that will be applied to support the introduction of the new brand.

APPLICATION

Background

The first application of the methodology involved a new brand of an aerosol deodorant product introduced by a competitor of the firm which sponsored the present work. Annual sales (at retail) for the product class in the United States amount to almost a half billion dollars and approximately a score of national brands were already being marketed before the emergence of the new brand. However, the two leading established brands held nearly half the market and the next five largest brands accounted for another 35% of the total product category volume. The new brand had been developed carefully and the basis of its positioning strategy was a straightforward but powerful claim of superior performance on an important attribute. The appearance of the new brand in a test market was regarded by management in this field as a competitive event of major importance.

Application of the system to this problem situation began after the new brand had been in test market in a midwestern city for eight months. Thus, carrying out the field work in a city *different* from that where the test market was underway afforded an opportunity to perform a test of the system's predictive ability in a relatively short period of time.

The design and conduct of the data collection corresponded to the methods and procedures summarized in Table 1. Two hundred and ninety-nine respondents were interviewed in a suburban shopping center of a city separate from, but similar to, the site of the test market then in progress. Quota sampling was used to obtain the desired representation of demographic characteristics and usage habits among persons interviewed. Respondents were shown television commercials for the five leading established brands plus one for the new brand. After giving his or her reaction to the commercials on a small set of rating scales, each respondent entered the simulated store with a coupon worth $2.00 in cash. Prices were set to be equal to the average of those prevailing in discount stores in the area at that time. Almost 75% of the sample bought one or another of the brands available which included the new one. Those who did not purchase the new brand were given a free sample as they left the store. Post-usage interviews were conducted by telephone three weeks later. Because the product is one typically used daily, this period was long enough for respondents to accumulate considerable usage experience with the new brand. Two-thirds of the original sample were reinterviewed and had been using the new brand.

Results

At several points in the foregoing discussions of measurement and estimation, data from the deodorant study were used to illustrate methods and results. Here attention is focused on the main predictions obtained from the models. Table 4 summarizes the inputs for the two models.

From the preference model, the average post-trial purchase probability for the new brand ($\Sigma L_i(t)/N$) was estimated to be .32. Estimates for the relevant set proportion ($E(t)$) were obtained by translating the $10 million annual national advertising spending rate estimated for the new brand into expected levels of advertising and brand awareness and then using these

[4]Both regressions are based on 18 observations. R^2 and SEE denote the coefficient of determination and the standard error of estimate, respectively. The figures in parentheses are the t statistics for the regression coefficients.

Table 4
MODEL INPUTS

	Quantity	Initial estimates	Adjusted estimates	Revised after test market
F	Conditional first purchase	.16	—	—
K	Brand awareness	.80	—	.83
D	Availability	.90	—	—
C	Sample coverage	.40	—	—
U	Sample usage	.75	—	.67
$R(k,t)$	Repeat rate	.11	.20	—
$R(t,t)$		.41	.55	—
Trial repeat model				
T	Cumulative trial	.381	—	.356
S	Repeat share	.157	.308	—
$M(t)$	Predicted market share (%)	6.0	11.7	11.0
Preference model				
$E(t)$	Relevant set proportion	.383~.445	—	.33
$\Sigma L_i(t)/N$	Avg. purchase prob.	.32	—	—
$M(t)$	Predicted market share (%)	12.3~14.2	—	10.6

values in the cross-sectional regression equations (22 and 23) for established brands to obtain predictions of the evoking proportion. This process yielded predictions for $E(t)$ of .383 (from the advertising awareness equation) and .445 (from the brand awareness equation). When combined in equation 5 with the estimate of .32 for the average post-trial purchase probability, these values of the relevant set parameter led to predicted market shares for the new brand of 12.3% and 14.2%, respectively.

The share prediction initially calculated from the trial-repeat model was much lower than the values obtained from the preference model. Although the value of the conditional probability of first purchase (F) estimated from the observed purchase rate of the new brand in the laboratory store was only .16, it was expected that a considerable amount of trial usage would be effected by a very extensive sampling program. When the introductory marketing plan was translated into the quantities specified in equation 14 it yielded a predicted value of .381 for the ultimate cumulative trial rate (T). This level of trial was consistent with the values of the evoking proportion (.38 to .445) estimated for the preference model. However, the repurchase inputs derived from the post-usage survey when applied to equation 15 led to an estimate of only .157 for the ultimate repeat purchase rate (S). This repeat (S) estimate plus the trial (T) level of .381 gave a predicted share of 6.0% for the new brand. The latter share was about half the 12.3–14.2% level predicted by the preference model, which used an average post-trial purchase probability estimate ($\Sigma \hat{L}_i(t)/N = .32$) roughly double the magnitude of the ultimate repeat rate quantity ($\hat{S} = .157$) estimated for the trial-repeat model.

Accounting for this marked discrepancy in the two repurchase estimates and hence the market share predictions was problematical. Efforts to uncover the source of the difficulty finally suggested a plausible diagnosis related to the measurements of the components of the repeat purchase rate, S. Recall from the discussion of the trial-repeat model that $R(t,t)$ is estimated by the proportion of respondents who make a "mail order repurchase" of the new brand when given the opportunity to do so in the post-usage survey. In this initial application, the parameter $R(k,t)$ was estimated from responses to a buying intentions scale rather than in the manner described heretofore which subsequently was adopted. As a consequence of these procedures, the estimate of the overall repeat rate S did not reflect any influence in in-store promotion or other external sources of reinforcement. However, such effects are implicitly represented in the calibration of the preference-purchase probability model. Furthermore, there was reason to believe that repurchase intentions expressed immediately after rejecting an opportunity to make a mail order purchase of the brand might be understated because respondents wished to discourage further solicitations. The influence of in-store promotion was known to be very important in this product category generally and the manufacturer of the new brand in particular has a reputation for using in-store activities aggressively as a means of stimulating repeat purchasing. For these reasons an upward adjustment of the observed levels of repurchase intentions appeared justifiable and the two repeat probabilities were raised judgmentally: $R(k,t)$ from .11 to .20 and $R(t,t)$, from .42 to .55. These modifications changed the ultimate repeat rate (S) from .157 to .308 and thereby raised the market share predicted by the trial-repeat model to 11.7% which is very close to the lower end of the 12.3–14.2% range obtained from the preference model.

The share prediction finally presented to management was the midpoint of the 12.3–14.2% range predicted by the preference model or 13.3%—a reflection of the greater confidence placed in the results obtained from the preference model in comparison with the trial-repeat model in this situation. The share observed in the test market, 12 months after launch, was 10.4%. The prediction exercise was carried out by the model building team while the new brand was in test market but before their exposure to any specific feedback or measurements on its early performance. As explained in the discussion of the models, the levels of certain marketing mix control variables that will persist in a test market must be specified in advance in order to develop predictions from the models. Here, the management group sponsoring this work had to supply these inputs for a competitor's brand rather than their own, and thus precise prior information was not available. In the course of reviewing the test market results, some significant differences were uncovered between the assumptions about the new brand's marketing plan that had been used in developing the predictions from the models and what actually had taken place in the test market. Taking account of the advertising and sampling programs which had in fact been used in the test market implied changes in the parameter estimates as indicated in the last column of Table 4 and, as expected, would have improved the accuracy of the market share prediction generated by the preference model. Whereas the difference between the share initially predicted and that observed in the test market was $13.3 - 10.4 = 2.9$ share points, the "revised deviation" or difference between the revised *ex post* prediction and the observed share was only $10.6 - 10.4 = 0.2$ share point.

Discussion

The foregoing discussion of the first application illustrates how features of the system and understanding of its capabilities and limitations have evolved. As additional applications have taken place, the adaptability of the procedures has been tested and certain modifications introduced to cope with new problems and to effect improvements. In the first study, the preference and trial-repeat models produced very different market share predictions and judgment had to be exercised to reconcile the discrepancies and arrive at a final prediction. After this experience, the change in the method of estimating the $R(k, t)$ parameter was adopted, but a completely satisfying explanation of the discrepancy has never been found. The practice of using both models has been continued and in more than 30 subsequent applications differences of the magnitude of those that arose in the first study have not been encountered. A Monté Carlo analysis performed in the trial-repeat model gave an estimate of 1.6 share points for the standard deviation of the model's market share predictions. This figure provides a rough basis for assessing disparities in the predictions given by the two models. If the differences appear to be within the bounds of sampling fluctuations, a simple average of the two outputs is used as the share prediction. When more substantial discrepancies occur, they must be interpreted and so judgment, guided by an examination of the diagnostic information obtained at several points in the measurement process outlined in Table 1, ultimately plays a role in deciding which results should be relied upon to obtain a final share prediction.

Available external validation data are not sufficient to allow clear discrimination between the two models and their measurement inputs. The time lags, attrition, and other exigencies normally encountered in the development of new packaged goods have made for a slow accumulation of opportunities for acquiring validation information. No tightly controlled tests of the present system's predictive accuracy have been performed. Cases in which new products have been subjected to both ASSESSOR *and* test market evaluations provide a basis for an early but only partial assessment of the quality of predictions generated. Of the approximately 30 new packaged goods studied with ASSESSOR to date, test marketing has been completed for nine and their final test market shares are known. Table 5 is a summary of results for these cases. The products are listed in the chronological order in which they were studied, beginning with the first application to the deodorant product which is included for completeness. Note that the ASSESSOR studies for the first three products were performed while their test markets were in progress and so are labelled "concurrent." These three applications were performed when the system first was developed and were conducted in this manner at the request of firms seeking information that would enable them to make an early evaluation of the system's predictive capability. In each of these nine cases, the ASSESSOR investigation was carried out in a city different from that used for the test market.

Table 5 shows the differences in the *initial* share predictions given by the preferences and trial-repeat models, *before* any reconciliation—i.e., the predictions based on planned or assumed test market programs. Except in the first application to the deodorant product, the discrepancies did not exceed one share point. For all nine products, including the first, the absolute average deviation was 1.2 share points which indicates that the predictions obtained from the two models generally have been in close agreement.

Also presented in Table 5 are the observed test market shares and the final share predictions made *after* comparing and where necessary reconciling judgmentally the separate predictions derived from the two models, but *before* test market results were known. Hence, these predictions do *not* reflect any *ex post* adjustments made to account for differences between planned and actual or implemented levels of marketing effort used in the test markets.

As can be seen from the table, the deviations

Table 5
PREDICTED AND OBSERVED MARKET SHARES

Product	Timing of pre-test in relation to test market: Concurrent	Before	Difference in share predictions of preference and trial-repeat models[a]	Market share (%): Predicted	Observed	Deviation[b]
Deodorant	√		+7.3	13.3	10.4	+2.9
Antacid	√		−0.9	9.6	10.5	−0.9
Laundry ingredient	√		+0.1	1.8	1.8[c] 2.0	−0.1
Household cleanser		√	−0.4	12.0	12.5	−0.5
Shampoo		√	+0.7	3.0	3.2	−0.2
Dishwashing ingredient		√	−0.2	9.3	8.5	+0.8
Pain reliever		√	+1.0	3.0	2.0	+1.0
Fruit drink		√	−0.2	4.9	5.0	−0.1
Cereal		√	+0.1	6.0	4.4	+1.6
Average (Absolute)			1.2	7.0	6.5	0.9

[a]Market share prediction obtained from the preference model minus that obtained from the trial-repeat model.
[b]Predicted minus observed market shares.
[c]Shares observed in two test market cities. The "observed" share used to calculate the "deviation" for this product was the mean of these two figures.

between the original predictions and the observed shares generally have been small, their absolute average being slightly less than one share point. However, the deviations in some instances appear more substantial when viewed as a percentage of the observed share, ranging from a low of 2% in the case of the fruit drink to a high of 50% for the pain reliever. As was noted for the deodorant application, seldom will the marketing mix program assumed in developing a prediction before the test market correspond exactly to that which is actually implemented later. Not surprisingly then, it has also been found for several of the subsequent applications that *ex post* predictions based on more precise knowledge of the marketing efforts expended in the test markets deviate less from the observed shares than do the original predictions shown in Table 5.

These results are reported in the spirit of revealing what is known about the accuracy of predictions developed through use of the system, but clearly they do not constitute a true predictive test. Though all the applications completed to date for which test market shares are available have been included, these cases are few in number and did not arise in a planned or prespecified manner. The lack of uniformity and precision associated with the observed test market shares themselves also deserves emphasis. The figures were obtained from several firms for the particular products whose investigation they had sponsored. Thus, the observed shares originated from several different sources, using a variety of methods. No claim can be made that the conditions of equivalence and independence have been met that enable unequivocal inferences about external validity to be drawn from comparisons of predicted and observed events.

The adequacy of the model's predictive ability must be evaluated in relation to how the model is used. At the pre-test-market stage, the manager is most interested in knowing whether he has a "winner." Will the brand earn a substantial share of the market? The second issue of concern to managers is how to improve the product's performance. The system's diagnostic capabilities and ability to make conditional forecasts for strategic changes can aid the manager in this task. Finally, the manager wants to know whether to drop the product, go to test market, or go national. If the predicted share is low and feasible changes in the marketing plan do not have potential to improve share substantially, dropping the product would be appropriate. If the share is good, either going to test or national introduction would be possible. The model proposed here does not answer this question. The manager could consider going national if the share is very high, investment is small, and competitive imitation is a danger. Usually, the product would go to test market. However, the test market would be oriented toward finding improvements in the marketing strategy rather than determining whether the product can attain an adequate market share. In this environment, the test market would be designed to place emphasis on measurement of response to marketing variables rather than determining share. The use of test market analysis models [e.g., 66] would be appropriate to process such data. If the test market confirms pre-test share estimates, the product could be introduced.

CONCLUSIONS

The authors have described a set of models and measurement procedures intended for use in evaluating new packaged goods at that stage in their development where management is faced with the decision of whether or not to place them in test markets. The approach taken to this problem is to merge relevant behavioral and management science concepts and methods. The results obtained from the initial applications have been sufficiently encouraging to suggest that the kind of methodology discussed can be a useful addition to the growing body of decision-support technology now available and being applied to the problems of managing new product development in the packaged goods field.

The system described is intended to aid management in evaluating a new packaged good brand at a particular point in the developmental process and it is important to recognize where the system can be expected to prove useful and where it may not. Experience gained from applications of the system as well as the nature of the models and measurement methodology itself suggest at least three factors or conditions as being necessary for obtaining satisfactory results. First, the applicability of the system is limited to situations in which the new brand seeks to penetrate a product category well-defined in terms of the nature and closeness of substitutes. Cases in which a very novel or innovative offering effectively creates a new product category cannot be handled by these methods. Second, the assumption that the usage/purchase rate for the new brand will be the same as that for the established brands must be tenable. Only limited means of coping with departures from this condition are available. A third restriction is that consumption and learning must occur at rates such that preferences for the new brand stabilize in a relatively short period. For products which are used infrequently or which require long periods of usage before benefits/satisfaction can be realized, it would not be feasible to measure post-usage preferences by the means described.

The development and evaluation of the system is an ongoing process. Additional tests bearing on the general issue of predictive validity will be possible in the future as test market data accumulate for products previously evaluated by this methodology. Future work will be undertaken to extend the range of new product situations to which the system can be applied.

REFERENCES

1. Achenbaum, Alvin A. "Market Testing: Using the Marketplace as a Laboratory," *in* Robert Ferber, ed., *Handbook of Marketing Research*. New York: McGraw-Hill Book Company, 1974, 4-31-54.
2. Ahl, David H. "New Product Forecasting Using Consumer Panels," *Journal of Marketing Research*, 7 (May 1970), 159–67.
3. Allaire, Yvan. "The Measurement of Heterogeneous Semantic, Perceptual, and Preference Structures," unpublished Ph.D. thesis, Massachusetts Institute of Technology, 1973.
4. ———, Gilles Laurent, and Alvin J. Silk. "The Goodness of Fit of Constant Sum Preference Scaling," in preparation.
5. Anonymous. "The Rebuilding Job at General Foods," *Business Week*, No. 2294 (August 25, 1973), 48–55.
6. Assmus, Gert. "NEWPROD: The Design and Implementation of a New Product Model," *Journal of Marketing*, 39 (January 1975), 16–23.
7. Axelrod, Joel N. "Attitude Measures that Predict Purchase," *Journal of Advertising Research*, 8 (March 1968), 3–18.
8. Ayal, Igal. "Trial and Repeat Buying—The Case for Separate Consideration," *Journal of the Academy of Marketing Science*, 2 (Summer 1974), 413–31.
9. Bass, Frank M. "The Theory of Stochastic Preference and Brand Switching," *Journal of Marketing Research*, 11 (February 1974), 1–20.
10. ———. "Analytical Approaches in the Study of Consumer Purchase Behavior and Brand Choice," Paper No. 520, Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Industrial Administration, Purdue University, July 1975.
11. ———, Edgar A. Pessemier, and Donald R. Lehmann. "An Experimental Study of Relationships Between Attitudes, Brand Preference, and Choice," *Behavioral Science*, 17 (November 1972), 532–41.
12. ———, Abel Jeuland, and Gordon P. Wright. "Equilibrium Stochastic Choice and Market Penetration Theories: Derivations and Comparisons," *Management Science*, 22 (June 1976), 1051–63.
13. Ben-Akiva, Moshe. "Program for Maximum Likelihood Estimation of the Multinomial Logit Model," unpublished paper, Transportation Systems Division, Department of Civil Engineering, Massachusetts Institute of Technology, January 1973.
14. Bird, Michael and A. S. C. Ehrenberg. "Non-Awareness and Non-Usage," *Journal of Advertising Research*, 6 (December 1966), 4–8.
15. Burger, Philip C. "COMP: A New Product Forecasting System," Working Paper 123-72, Graduate School of Management, Northwestern University, 1972.
16. Buzzell, Robert D. and Robert E. M. Nourse. *Product Innovation in Food Processing, 1954–1964*. Boston, Massachusetts: Division of Research, Graduate School of Business Administration, Harvard University, 1967.
17. Cadbury, N. D. "When, Where, and How to Test Market," *Harvard Business Review*, 53 (May–June 1975), 96–105.
18. Campbell, Brian M. "The Existence of Evoked Set and Determinants of its Magnitude in Brand Choice Behavior," unpublished Ph.D. dissertation, Columbia University, 1969.
19. Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs in Research*. Chicago: Rand-McNally and Company, 1963.
20. Charlton, P., A. S. C. Ehrenberg, and B. Pymont. "Buyer Behavior Under Mini Test Conditions," *Journal of the Market Research Society*, 14 (July 1972), 171–83.
21. ——— and B. Pymont. "Evaluating Marketing Alterna-

tives," *Journal of the Market Research Society,* 17 (April 1975), 90–103.

22. Claycamp, Henry J. and Lucien E. Liddy. "Prediction of New Product Performance: An Analytical Approach," *Journal of Marketing Research,* 6 (November 1969), 414–20.
23. Coombs, Clyde H., Robyn M. Dawes, and Amos Tversky. *Mathematical Psychology.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.
24. Domencich, Thomas A. and Daniel McFadden. *Urban Travel Demand.* Amsterdam: North-Holland Publishing Co., 1975.
25. Ehrenberg, A. S. C. *Repeat-Buying.* Amsterdam: North-Holland Publishing Co., 1972.
26. ———. "Predicting the Performance of New Brands," *Journal of Advertising Research,* 11 (December 1971), 3–10.
27. ——— and G. J. Goodhardt. "Repeat Buying of a New Brand," *British Journal of Marketing,* 2 (Autumn 1968), 200–5.
28. Eskin, Gerald J. and John Malec. "A Model for Estimating Sales Potential Prior to the Test Market," in *Proceedings* 1976 Fall Educators' Conference, Series No. 39. Chicago: American Marketing Association, 1976, 230–3.
29. Forecast Market Research, Ltd. *The Mini-Test Market.* London: 1976.
30. Ginter, James L. "An Experimental Investigation of Attitude Change and Choice of a New Brand," *Journal of Marketing Research,* 11 (February 1974), 30–40.
31. Green, Paul E. "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing,* 39 (January 1975), 24–31.
32. Haley, Russell I. "We Shot an Arrowhead into the Air," *Proceedings,* 16th Annual Conference, Advertising Research Foundation, New York, 1970, 25–30.
33. Hardin, D. K. "Changing Test Market Technology in the U.S.A.," in *Marketing in a Changing World: The Role of Market Research,* ESOMAR Proceedings, 1972 Congress, Cannes, France. Amsterdam, The Netherlands: European Society of Opinion and Marketing Research, 315–39.
34. Hatch, Thomas E. and Glen L. Urban. "New Products—Minimizing Risk and Maximizing Creativity," Working Paper 752-74, Sloan School of Management, Massachusetts Institute of Technology, November 1974.
35. Hauser, John R. "A Normative Methodology for Predicting Consumer Response to Design Decisions: Issues, Models, Theory and Use," unpublished Ph.D. thesis, Massachusetts Institute of Technology, 1975.
36. ———. "The Information Measure Provides a Mathematically Rigorous and Intuitively Appealing Predictive Test for Probabilistic Demand Models," Working Paper, Transportation Center, Northwestern University, March 1976.
37. Howard, John A. and Jagdish N. Sheth. *The Theory of Buyer Behavior.* New York: John Wiley and Sons, Inc., 1969.
38. Jacoby, Jacob. "A Model of Multi-Brand Loyalty," *Journal of Advertising Research,* 11 (June 1971), 25–31.
39. Kendall, Maurice G. *Rank Correlation Methods.* London: Griffin, 1955.
40. Kesavan, R. and V. Srinivasan. "Bias in Stochastic Binary Models of Brand Choice," in *Proceedings* 1976 Fall Educators' Conference, Series No. 39, Chicago: American Marketing Association, 115–9.
41. Kotler, Philip. *Marketing Decision Making.* New York: Holt, Rinehart and Winston, 1971.
42. Lipstein, Benjamin. "The Design of Test Marketing Experiments," *Journal of Advertising Research,* 5 (December 1965), 2–7.
43. Luce, R. Duncan. *Individual Choice Behavior.* New York: John Wiley and Sons, Inc., 1969.
44. Massy, William F., Ronald E. Frank, and Thomas M. Lodahl. *Purchasing Behavior and Personal Attributes.* Philadelphia: University of Pennsylvania Press, 1968.
45. ———, David B. Montgomery, and Donald F. Morrison. *Stochastic Models of Buyer Behavior.* Cambridge, Massachusetts: MIT Press, 1970.
46. McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior," *in* Paul Zarembka, ed., *Frontiers in Econometrics.* New York: Academic Press, 1970, 105–42.
47. McGuire, E. Patrick. *Evaluating New Product Proposals.* New York: The Conference Board, Report No. 604, 1973.
48. Narayana, Chem L. and Rom J. Markin. "Consumer Behavior and Product Performance: An Alternative Conceptualization," *Journal of Marketing,* 39 (October 1975), 1–6.
49. National Industrial Conference Board. *Market Testing Consumer Products.* New York: National Industrial Conference Board, Experiences in Marketing Management, No. 12, 1967.
50. Nielsen Marketing Service. "New Brand or Superbrand?" *The Nielsen Researcher,* 29, No. 5 (1971), 4–10.
51. ———. "To Test or Not to Test," *The Nielsen Researcher,* 30, No. 4 (1972), 3–8.
52. Parfitt, J. H. and B. J. K. Collins. "Use of Consumer Panels for Brand Share Prediction," *Journal of Marketing Research,* 5 (May 1968), 131–46.
53. Pessemier, Edgar A. "Market Structure Analysis of New Product and Market Opportunities," *Journal of Contemporary Business,* 4 (Spring 1975), 35–65.
54. ——— and Richard D. Teach. "A Single Subject Scaling Model Using Judged Distances Between Pairs of Stimuli," Paper No. 143, Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Industrial Administration, Purdue University, August 1966.
55. ———, Philip Burger, Richard Teach, and Douglas Tigert. "Using Laboratory Brand Preference Scales to Predict Consumer Brand Purchases," *Management Science,* 17 (February 1971), B-371-385.
56. ——— and William L. Wilkie. "Multi-Attribute Choice Theory—A Review and Analysis," Paper No. 372, Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Industrial Administration, Purdue University, September 1972.
57. Pymont, B. C. "The Development and Application of a New Micro Market Testing Technique," in *The Practical Applications of Market Research,* ESOMAR Papers, 1970 Congress, Barcelona, Spain. Amsterdam, The Netherlands: European Society of Opinion and Marketing Research, 201–21.
58. ———, D. Reag, and P. G. M. Standen. "Towards The Elimination of Risk From Investment in New Products:

Experience with Micro-Market Testing," paper presented at the 1976 ESOMAR Congress, Venice, Italy, September 1976.
59. Rao, Vithala. "The Salience of Price in the Perception and Evaluation of Product Quality: A Multidimensional Measurement Model and Experimental Test," unpublished Ph.D. thesis, University of Pennsylvania, 1970.
60. Shocker, Allan D. and V. Srinivasan. "A Consumer-Based Methodology for the Identification of New Product Ideas," *Management Science,* 20 (February 1974), 921-37.
61. Stanton, Frank. "What is Wrong with Test Marketing?", *Journal of Marketing,* 31 (April 1967), 43-7.
62. Stefflre, Volney. "Market Structure Studies: New Products for Old Markets and New Markets (Foreign) for Old Products," *in* Frank M. Bass, Charles W. King, and Edgar A. Pessemier, eds., *Applications of the Sciences in Marketing.* New York: John Wiley and Sons, Inc., 1968, 251-68.
63. Stevens, S. S. "Ratio Scales of Opinion," *in* Dean Whitla, ed., *Handbook of Measurement and Assessment in Behavioral Sciences.* New York: Addison-Wesley Publishing Company, 1968, 171-99.
64. Tauber, Edward M. "Forecasting Sales Prior to Test Market," *Journal of Marketing,* 41 (January 1977), 80-4.
65. Torgerson, Warren S. *Theory and Method of Scaling.* New York: John Wiley and Sons, Inc., 1958.
66. Urban, Glen L. "SPRINTER Mod III: A Model for Analysis of New Frequently Purchased Consumer Products," *Operations Research,* 18 (September-October 1970), 805-53.
67. ———. "PERCEPTOR: A Model for Product Positioning," *Management Science,* 21 (April 1975), 858-71.
68. ——— and Richard Karash. "Evolutionary Model Building," *Journal of Marketing Research,* 8 (February 1971), 62-6.
69. Wachsler, Robert A., Lewis G. Pringle, and Edward I. Brody. "NEWS: A Systematic Methodology for Diagnosing New Product Marketing Plans and Developing Actionable Recommendations," unpublished paper, Research Department, Batten, Barton, Durstine and Osborn, Inc., New York, February 1972.
70. Wierenga, B. *An Investigation of Brand Choice Processes.* Rotterdam: University of Rotterdam Press, 1974.
71. Wilkie, William L. and Edgar A. Pessemier. "Issues in Marketing's Use of Multi-Attribute Attitude Models," *Journal of Marketing Research,* 10 (November 1973), 428-41.
72. Wind, Yoram. "A New Procedure for Concept Testing," *Journal of Marketing,* 37 (October 1973), 2-11.
73. Winter, Frederick W. "A Laboratory Experiment of Individual Attitude Response to Advertising Exposure," *Journal of Marketing Research,* 10 (May 1973), 130-40.

LINKED CITATIONS

You have printed the following article:

Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology
Alvin J. Silk; Glen L. Urban
Journal of Marketing Research, Vol. 15, No. 2. (May, 1978), pp. 171-191.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197805%2915%3A2%3C171%3APEONPG%3E2.0.CO%3B2-P

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

[1] **Consumer Behavior and Product Performance: An Alternative Conceptualization**
Chem L. Narayana; Rom J. Markin
Journal of Marketing, Vol. 39, No. 4. (Oct., 1975), pp. 1-6.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197510%2939%3A4%3C1%3ACBAPPA%3E2.0.CO%3B2-X

[2] **The Theory of Stochastic Preference and Brand Switching**
Frank M. Bass
Journal of Marketing Research, Vol. 11, No. 1. (Feb., 1974), pp. 1-20.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197402%2911%3A1%3C1%3ATTOSPA%3E2.0.CO%3B2-R

References

[2] **New Product Forecasting Using Consumer Panels**
David H. Ahl
Journal of Marketing Research, Vol. 7, No. 2. (May, 1970), pp. 160-167.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197005%297%3A2%3C160%3ANPFUCP%3E2.0.CO%3B2-%23

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

[6] **Newprod: The Design and Implementation of a New Product Model**
Gert Assmus
Journal of Marketing, Vol. 39, No. 1. (Jan., 1975), pp. 16-23.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197501%2939%3A1%3C16%3ANTDAIO%3E2.0.CO%3B2-M

[9] **The Theory of Stochastic Preference and Brand Switching**
Frank M. Bass
Journal of Marketing Research, Vol. 11, No. 1. (Feb., 1974), pp. 1-20.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197402%2911%3A1%3C1%3ATTOSPA%3E2.0.CO%3B2-R

[12] **Equilibrium Stochastic Choice and Market Penetration Theories: Derivations and Comparisons**
Frank M. Bass; Abel Jeuland; Gordon P. Wright
Management Science, Vol. 22, No. 10. (Jun., 1976), pp. 1051-1063.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197606%2922%3A10%3C1051%3AESCAMP%3E2.0.CO%3B2-N

[22] **Prediction of New Product Performance: An Analytical Approach**
Henry J. Claycamp; Lucien E. Liddy
Journal of Marketing Research, Vol. 6, No. 4. (Nov., 1969), pp. 414-420.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28196911%296%3A4%3C414%3APONPPA%3E2.0.CO%3B2-4

[30] **An Experimental Investigation of Attitude Change and Choice of a New Brand**
James L. Ginter
Journal of Marketing Research, Vol. 11, No. 1. (Feb., 1974), pp. 30-40.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197402%2911%3A1%3C30%3AAEIOAC%3E2.0.CO%3B2-Y

[31] **Marketing Applications of MDS: Assessment and Outlook**
Paul E. Green
Journal of Marketing, Vol. 39, No. 1. (Jan., 1975), pp. 24-31.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197501%2939%3A1%3C24%3AMAOMAA%3E2.0.CO%3B2-8

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

[48] **Consumer Behavior and Product Performance: An Alternative Conceptualization**
Chem L. Narayana; Rom J. Markin
Journal of Marketing, Vol. 39, No. 4. (Oct., 1975), pp. 1-6.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197510%2939%3A4%3C1%3ACBAPPA%3E2.0.CO%3B2-X

[52] **Use of Consumer Panels for Brand-Share Prediction**
J. H. Parfitt; B. J. K. Collins
Journal of Marketing Research, Vol. 5, No. 2. (May, 1968), pp. 131-145.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28196805%295%3A2%3C131%3AUOCPFB%3E2.0.CO%3B2-K

[55] **Using Laboratory Brand Peference Scales to Predict Consumer Brand Purchases**
Edgar Pessemier; Philip Burger; Richard Teach; Douglas Tigert
Management Science, Vol. 17, No. 6, Application Series. (Feb., 1971), pp. B371-B385.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197102%2917%3A6%3CB371%3AULBPST%3E2.0.CO%3B2-E

[60] **A Consumer-Based Methodology for the Identification of New Product Ideas**
Allan D. Shocker; V. Srinivasan
Management Science, Vol. 20, No. 6, Application Series. (Feb., 1974), pp. 921-937.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197402%2920%3A6%3C921%3AACMFTI%3E2.0.CO%3B2-A

[61] **What Is Wrong with Test Marketing?**
Frank Stanton
Journal of Marketing, Vol. 31, No. 2. (Apr., 1967), pp. 43-47.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28196704%2931%3A2%3C43%3AWIWWTM%3E2.0.CO%3B2-6

[64] **Forecasting Sales Prior to Test Market**
Edward M. Tauber
Journal of Marketing, Vol. 41, No. 1. (Jan., 1977), pp. 80-84.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197701%2941%3A1%3C80%3AFSPTTM%3E2.0.CO%3B2-2

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

[66] **Sprinter Mod III: A Model for the Analysis of New Frequently Purchased Consumer Products**
Glen L. Urban
Operations Research, Vol. 18, No. 5. (Sep. - Oct., 1970), pp. 805-854.
Stable URL:
http://links.jstor.org/sici?sici=0030-364X%28197009%2F10%2918%3A5%3C805%3ASMIAMF%3E2.0.CO%3B2-P

[67] **Perceptor: A Model for Product Positioning**
Glen L. Urban
Management Science, Vol. 21, No. 8, Application Series. (Apr., 1975), pp. 858-871.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197504%2921%3A8%3C858%3APAMFPP%3E2.0.CO%3B2-V

[68] **Evolutionary Model Building**
Glen L. Urban; Richard Karash
Journal of Marketing Research, Vol. 8, No. 1. (Feb., 1971), pp. 62-66.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197102%298%3A1%3C62%3AEMB%3E2.0.CO%3B2-E

[71] **Issues in Marketing's Use of Multi-Attribute Attitude Models**
William L. Wilkie; Edgar A. Pessemier
Journal of Marketing Research, Vol. 10, No. 4. (Nov., 1973), pp. 428-441.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197311%2910%3A4%3C428%3AIIMUOM%3E2.0.CO%3B2-I

[72] **A New Procedure for Concept Evaluation**
Yoram Wind
Journal of Marketing, Vol. 37, No. 4. (Oct., 1973), pp. 2-11.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197310%2937%3A4%3C2%3AANPFCE%3E2.0.CO%3B2-M

[73] **A Laboratory Experiment of Individual Attitude Response to Advertising Exposure**
Frederick W. Winter
Journal of Marketing Research, Vol. 10, No. 2. (May, 1973), pp. 130-140.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197305%2910%3A2%3C130%3AALEOIA%3E2.0.CO%3B2-A

NOTE: *The reference numbering from the original has been maintained in this citation list.*