



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### A Logit Model of Brand Choice Calibrated on Scanner Data

Peter M. Guadagni, John D. C. Little,

To cite this article:

Peter M. Guadagni, John D. C. Little, (2008) A Logit Model of Brand Choice Calibrated on Scanner Data. Marketing Science 27(1):29-48. <https://doi.org/10.1287/mksc.1070.0331>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2008, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# A Logit Model of Brand Choice Calibrated on Scanner Data

Peter M. Guadagni

Bonanza Street Books, Walnut Creek, California 94596, guadagni@alum.mit.edu

John D. C. Little

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, jlittle@mit.edu

A multinomial logit model of brand choice, calibrated on 32 weeks of purchases of regular ground coffee by 100 households, shows high statistical significance for the explanatory variables of brand loyalty, size loyalty, presence/absence of store promotion, regular shelf price and promotional price cut. The model is parsimonious in that the coefficients of these variables are modeled to be the same for all coffee brand-sizes. The calibrated model predicts remarkably well the share of purchases by brand-size in a hold-out sample of 100 households over the 32-week calibration period and a subsequent 20-week forecast period. The success of the model is attributed in part to the level of detail and completeness of the household panel data employed, which has been collected through optical scanning of the Universal Product Code in supermarkets.

Three short-term market response measures are calculated from the model: regular (depromoted) price elasticity of share, percent increase in share for a promotion with a median price cut, and promotional price cut elasticity of share. Response varies across brand-sizes in a systematic way with large share brand-sizes showing less response in percentage terms but greater in absolute terms. On the basis of the model a quantitative picture emerges of groups of loyal customers who are relatively insensitive to marketing actions and a pool of switchers who are quite sensitive.

This article was originally published in *Marketing Science*, Volume 2, Issue 3, pages 203–238, in 1983.

*Key words:* choice; logit; marketing-mix; scanners

*History:* This paper was received May 1982 and was with the authors for 2 revisions.

## 1. Introduction

Manufacturers and retailers wish to understand how price, promotion and other marketing variables affect the sales and shares of the products they sell. Such information is the raw material for marketing mix decisions. Although first priority goes to determining how a product's variables affect its own sales, marketing managers increasingly would like to learn more about product interactions within a category. For example, a manufacturer would like to know whether promoting one product takes away sales and share from others in the same line. Similarly, retailers are acutely aware that a price cut on an item may increase its sales at the expense of a related item somewhere else in the store.

To understand such issues we need to model whole product categories. This task can be partitioned into determining, first, the effects of marketing variables on share and, then, effects on total category sales. In this paper, we address only the question of share, or, to be more precise, we examine the effect of marketing variables on customer choice among product alternatives. From choices we deduce share.

Many models of choice have been proposed. In fact, theoretical developments seem to be ahead of empir-

ical testing, at least insofar as live marketing practice is concerned. Fortunately, the automatic recording of purchases at the point of sale opens up new opportunities for model appraisal. Data now being collected by optical scanning of the Universal Product Code (UPC) in supermarkets should permit a careful examination of various customer choice models in the case of grocery products. We shall here use scanner-collected data on coffee purchases to calibrate a multinomial logit choice model and examine both its scientific quality as a representation of customer behavior and its potential usefulness for marketing decisionmaking.

Scanner data usually come in two forms: *Store data* and *panel data*. *Store data* provide individual item (UPC) sales and price by store by week. In addition, the companies supplying data may collect information on other store activities such as special display, coupon redemption, retail advertising, and shelf-space allocation. *Panel data* present histories of purchases for a sample of households. A cooperating household displays an identification card at checkout. The store clerk keys the household number into the cash register, thereby causing the purchase record to

be segregated and stored. Over time this creates a longitudinal customer history.

Scanner data have special advantages. They directly record the sales act of the individual customer at the item level, thereby avoiding the pipeline effects found in factory shipments and warehouse withdrawals. Initial cost is low since the data are a spinoff from the store's basic transaction process. However, subsequent processing and storage can be expensive. Scanner data are potentially deliverable to users with great speed, although how fast they will actually be delivered will depend on their time-value in decisionmaking. The data tend to be very accurate because they are part of the store's cash collection and accounting process. At present some purchases enter the cash register around the scanner, but, since store procedures determine the percentage of goods scanned, as more stores use the data for inventory control and other informational purposes, accuracy will further increase. Compared to diary panels, scanner measurement is relatively unobtrusive, bias-free and complete across products. It therefore represents a significant step forward in gathering information about products sold through supermarkets.

The greatest advantage of scanner data, however, is that they provide the competitive environment of the customer decision. Conventional diary data tell what the customer bought and its price but do not identify the other products, prices, and marketing activities impinging on the customer at the time of purchase. Similarly, standard warehouse withdrawal or store audit data for a geographic region do not describe the competitive situation within individual stores in the way scanner store data do. It is this rich, disaggregated detail that offers hope for new levels of customer and market understanding. Just as new measurements have driven advances in theory throughout the history of science, we can hope for progress here.

Current scanner data services fall in three broad categories: (1) *groups of stores* in single markets, (2) *national samples of stores*, and (3) *instrumented markets*. A *group of stores* in a single market offers opportunities for convenient in-store experiments using scanner-collected store data. When panels of store-loyal shoppers are added, the experimentation possibilities increase. Coupons, for example, can be validly tested for the first time, and models of individual customer behavior with respect to price and promotion become possible.

*National store samples* permit manufacturers to make generalizations that might be distrusted if made from a single market. Good random samples are still difficult to obtain because of the irregular geographic distribution of scanners but this will change. Empirical studies of store data can determine sales response

to promotion and price at the store level and, by looking across stores, information can be gathered on retailer's response to manufacturers' promotional offerings.

The most dramatic new service stimulated by scanners, however, is the *instrumented market*. This is a small to medium sized city with scanners in all major grocery stores. In addition, the city is pre-selected for high cable television coverage and split cable hardware is introduced so that household panels can be set up to receive different television advertisements. In-store observations can also be conducted, and so a remarkably complete picture of the shopper's marketing environment is possible: sales, prices, promotions, advertising (both newspaper and television), coupons, display and shelf-facings. The setting is ideal for a variety of testing: new products, advertising, store promotions, etc.

Much measurement and model building lies ahead using these various services. The present paper is one step. It will focus on a choice model applicable to household data collected in panels associated with individual stores or instrumented markets, supplemented by store data on the shopping environment. To indicate the task at hand, Figure 1 displays an example of scanner collected data. The market share of a major brand of coffee in a panel of 100 households shows great variation over a year's time both in overall trend and in specific peaks. We wish to understand and predict such behavior.

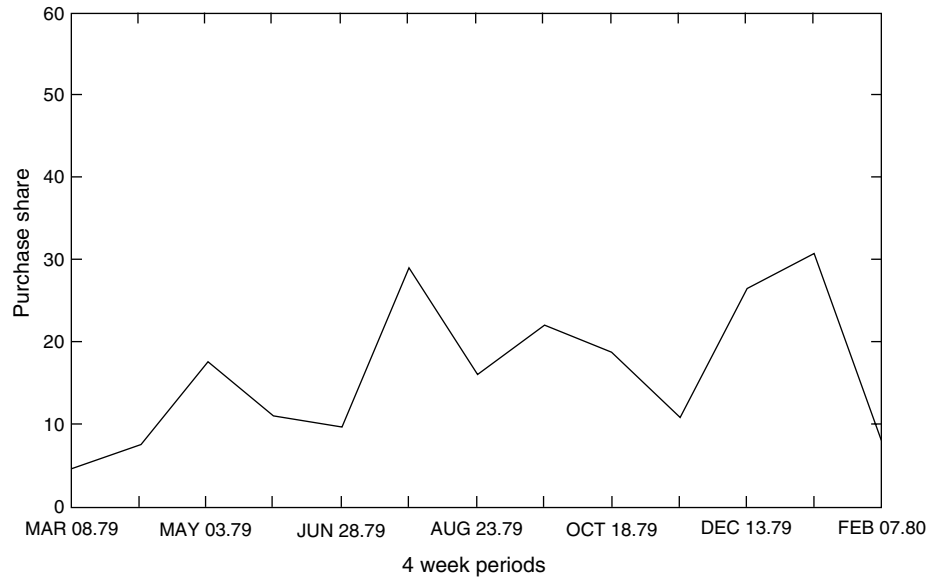
## 2. The Multinomial Logit Choice Model

The multinomial logit model computes the probability of choosing an alternative as a function of the attributes of all the alternatives available. The model has the appeal of being stochastic and yet admitting decision variables. Various authors have employed it in marketing. Hlavac and Little (1966) use a somewhat similar model to represent the probability that an automobile buyer purchases a car at a particular dealership. Silk and Urban (1978) imbed the logit in their pre-test-market evaluation process for new products. Punj and Staelin (1978) employ the model to describe students' choice of business schools. Gensch and Recker (1979) provide a general exposition and compare the fitting ability of the logit to that of regression for shoppers' choosing grocery stores. The logit has an even more extensive history of application in the field of transportation planning, particularly for predicting an individual's choice of mode of travel, e.g., car or bus (Domencich and McFadden 1975).

### 2.1. Axiomatic View

As a choice model, the multinomial logit permits an axiomatic derivation which we briefly outline. Consider an *individual*, *i*, confronted with a choice from

**Figure 1** The Share of Purchases of a Major Coffee Brand-Size as Recorded in a Panel of 100 Kansas City Households Shows Great Variability (Dates Shown Are the Starting Days of Four-Week Periods)



a set,  $S_i$ , of *alternatives*. In our setting the alternatives will be different products in a category. We suppose that:

(1) Alternative  $k \in S_i$  holds for the individual a preference or *utility*,

$$u_k = v_k + \epsilon_k, \quad (1)$$

where

$v_k$  = a deterministic component of  $i$ 's utility, to be calculated from observed variables, and

$\epsilon_k$  = a random component of  $i$ 's utility, varying from choice occasion to choice occasion, possibly as a result of unobserved variables.

(2) Confronted by the set of alternatives, individual  $i$  chooses the one with the highest utility on the occasion. I.e., the probability of choosing  $k$  is

$$p_k = P\{u_k \geq u_j, j \in S_i\}. \quad (2)$$

(3) The  $\epsilon_k, k \in S_i$ , are independently distributed random variables with a double exponential (Gumbel type II extreme value) distribution

$$P(\epsilon_k \leq \epsilon) = e^{-e^{-\epsilon}}, \quad -\infty < \epsilon < \infty. \quad (3)$$

This form of the distribution appears to fix the mean and variance of  $\epsilon$  quite arbitrarily, since (3) has a mean 0.575 and a variance 1.622, both dimension-less. A more general form would include a further location parameter and a scale parameter. However, any location parameter, even one dependent on  $k$ , can be absorbed into  $v_k$  without loss of generality and, since the scaling of utility is arbitrary, we can set it so that the variance of the  $\epsilon_k$  is the 1.622 value implied by (3).

Notice, however, that this procedure produces larger utility values in a model that explains more variance than in one that explains less. We shall observe this phenomenon in our empirical work.

Given assumptions (1)–(3), it can be shown (Theil 1969, McFadden 1974) that individual  $i$ 's choice probabilities have the remarkably simple form

$$p_k = e^{v_k} / \sum_{j \in S_i} e^{v_j}. \quad (4)$$

This expression is known as the multinomial logit.

We note two properties that will be used later. First, since (4) can be written

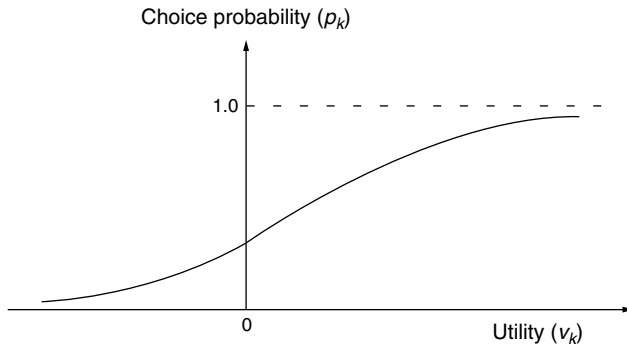
$$p_k = 1 / \sum_j e^{(v_j - v_k)},$$

it follows that utility is undetermined to the extent of an additive constant. Thus, for example, if a price variable has an inflationary trend that adds a constant to all alternatives,  $p_k$  will not be affected.

Second,  $p_k$  is S-shaped in  $v_k$  when other  $v_j$  are held constant. Therefore, as shown in Figure 2, very large or very small values of  $v_k$  make  $p_k$  flat and insensitive to changes in  $v_k$ .

We also note that, as the variance of the random component of utility goes to zero, the scaling of (3) pushes individual  $v_j$  (and any differences between pairs of  $v_j$ ) toward infinity. As a result, the largest  $v_k$  produces a  $p_k$  that goes to 1, while others go to 0, as we would wish.

In our case the individual choice-makers are households. We do not know whether their behavior satisfies the assumptions used to derive (4). However, the

**Figure 2** Choice Probability Is S-Shaped in Utility  $v_k$ 

concept of utility (or preference or attractiveness) as a latent variable and a choice probability that is some normalized function of that variable has much appeal and a long history (Luce 1959, Yellott 1977, McFadden 1981).

A chief complaint about (4) is that, if we add to the alternatives a new one essentially identical to some existing alternative, say the  $k$ th, the new alternative might reasonably be expected to split  $k$ 's probability and leave the others untouched, but, by (4), will instead reduce the probabilities of all alternatives. The issue is whether choices satisfy the assumption of "independence from irrelevant alternatives." (See Appendix 1 for further discussion.)

Various schemes have been proposed for overcoming the potential difficulty (McFadden 1981). Many of them involve arranging the alternatives into a hierarchy or tree structure that groups similar alternatives. Thus a tree representation of coffees might have all instant brands in one branch and all regular brands in another. We shall stay with the simple multinomial (4) adopting the pragmatic view that tests of the calibrated model will determine its quality. In setting up the application, however, we endeavor to avoid its known pitfalls.

## 2.2. Linear Utility

The deterministic component of a customer's utility for alternative  $k$  will be expressed as a linear function of observed variables, called the *attributes* of  $k$ . Some of these will be attributes of the product (e.g., price) and others may be attributes of the customer or the environment (e.g., income or store) that differentially favor one alternative over another for some reason. In general

$$v_k^i = \sum_{j \in T} b_{jk} x_{jk}^i \quad (5)$$

where

$x_{jk}^i$  = observed value of attribute  $j$  of alternative  $k$  for customer  $i$ ,

$b_{jk}$  = utility weight of attribute  $j$  of alternative  $k$ . We shall drop the superscript  $i$  when it is not required for clarity.

From a modeling point of view it is convenient to break the attributes into two classes.

(1)  $T_k = \{\text{attributes unique to alternative } k\}$ . A product may have features that other products do not have but which contribute to its utility. For such attributes  $j \in T_k$ , the coefficients may be denoted  $b_{jk}$  and variables by  $x_{jk}^i$ .

(2)  $T_C = \{\text{attributes common to all alternatives}\}$ . Although an attribute such as price might be assigned uniquely to each product, a model more parsimonious in the number of parameters would use price as a single attribute across all products. Then price would have the same coefficient for all alternatives. For such attributes  $j \in T_C$ , the coefficients may be denoted  $b_j$  and variables  $x_{jk}^i$ .

In specialized form (5) becomes

$$v_k^i = \sum_{j \in T_k} b_{jk} x_{jk}^i + \sum_{j \in T_C} b_j x_{jk}^i. \quad (6)$$

Although a linear form for utility is a natural place to start model building, we note that linearity for  $v_k$  still leaves the choice probability quite nonlinear in the observational variables,  $x_{jk}$ . The numerator of  $p_k$  can be written

$$e^{v_k} = \prod_{j \in T} e^{b_{jk} x_{jk}}$$

so that the model is, in an important sense, more multiplicative than additive.

## 2.3. Calibration

Equations (4) and (6) present the model. In practice we cannot observe either utilities or probabilities directly. Rather, we observe choices and attribute values. The data consist of a set of choice records for each individual:

$$y_k^i(n) = \begin{cases} 1 & \text{if customer } i \text{ chooses alternative } k \\ & \text{on the } n\text{th choice occasion,} \\ 0 & \text{otherwise,} \end{cases}$$

along with the values of the attributes of the alternatives on each choice occasion:

$$x_{jk}^i(n) = \text{value of attribute } j \text{ for product } k \text{ on } n\text{th purchase occasion for customer } i.$$

In our case a choice or observation is the purchase of a product by a customer on the occasion of buying within the product class. The  $b_{jk}$  and  $b_j$  are unknown constants to be determined by calibration. For calibration purposes, each attribute is thought of as having a complete set of data across alternatives for each observation, even though an attribute unique to a specific alternative does not appear in the utility expressions for other alternatives. To handle this situation, such an attribute is assigned a zero value for alternatives to which it is not relevant.

Calibration is done by maximum likelihood, using (4) and (6) to calculate the likelihood function. The actual program used is that of Manski and Ben-Akiva (Ben-Akiva 1973). Maximum likelihood parameter estimates are consistent and asymptotically efficient and normally distributed under very general conditions.

## 2.4. Quality of Fit

Measures of quality of fit and parameter estimation guide model specification and help appraise the success of the calibration.

(1) *t-values for coefficients*. The Manski–Ben-Akiva program generates (asymptotic) *t*-statistics for each coefficient in the calibrated model.

(2)  $U^2$  for model. Probabilistic models pose special difficulties in overall evaluation. Whereas regression models offer residuals and  $R^2$  as ready indicators of fit, a logit model predicts only probabilities which must then be compared to actual choices. Hauser (1978) discusses this and presents a set of statistics useful for evaluating such models. Working from the information measure of communication theory, he defines a measure  $U^2$ , in terms of calibrated model and prior probabilities of choice.  $U^2$  may be described as the fraction of uncertainty (entropy) empirically explained by the calibrated model relative to the prior distribution of choice probabilities. The prior distribution constitutes a null model that defines maximum entropy for the situation at hand. A model under test that produces the same entropy as the null model explains nothing new and has  $U^2 = 0$ . A model that explains everything gives perfect prediction and has  $U^2 = 1$ .

Hauser derives a useful result for the case where the prior probability of choice is independent of the choice occasion. (This will be true for the null model we shall use, namely, choice probabilities equal to market shares.) He shows that under these circumstances  $U^2$  equals McFadden's (1974) likelihood ratio index,  $\rho^2$ . Therefore, using McFadden's definition of  $\rho^2$ , we have

$$U^2 = \rho^2 = 1 - L(X)/L_0 \quad (7)$$

where  $L(X)$  is the log likelihood of the calibrated model with explanatory variables,  $X$ , and  $L_0$  is the log likelihood of the null model. Notice that, if  $L(X)$  of the calibrated model does not improve on  $L_0$ , then  $L(X) = L_0$  and  $U^2 = 0$ . However, if the model is perfect, i.e., the predicted probabilities are all zeros and ones and correct, then the likelihood equals 1,  $L(X) = 0$  and  $U^2 = 1$ . For the case that interests us, therefore, Hauser's and McFadden's quite differently motivated measures are equal and, since the Manski–Ben-Akiva program provides log likelihoods in the output, we can easily calculate  $U^2$  from (7).

$U^2$  is somewhat analogous to  $R^2$  in that they both have a range of 0 to 1 and indicate degree of variability explained.  $U^2$  tends, however, to have a lower value for an excellent fit.

(3) *Chi-squared tests of model significance*. If one model, say A, can be formulated as a restriction (subset) of the parameters of the tested model, say B, then  $L = 2 \log$  [likelihood ratio of model B to model A] is  $\chi^2$  distributed with degrees of freedom equal to the difference in degrees of freedom between model B and model A. See Hauser (1978). This test helps determine whether adding a parameter or set of parameters is worthwhile.

(4) *Aggregate share*. For a given population the average probability of choosing an alternative is the expected share of choices for that alternative. We can compare actual with expected share. Such comparisons plotted over time offer valuable visual representations of quality of fit.

## 3. Application to Packaged Goods

Market share is an aggregation of individual customer choices. If we can understand how and why households choose one product over another, we shall gain insight into the reasons for a product's success or failure. Scanner panel data and the logit model offer an opportunity for increasing our knowledge about packaged goods.

In applying the model, the alternatives will be products, but their exact level of aggregation and which ones to include in the relevant set are not necessarily obvious. Should different flavors, or colors, or sizes be treated as different products or lumped together? When product choice is hierarchical, we can induce similarity by defining alternatives at the same level in the hierarchy. For example, within the category of vegetables, the choice might be among fresh, frozen and canned. However, in studying brands, we might focus on competition within, say, canned peas. In the case of coffee, a natural grouping exists segregating instant and ground forms and another distinguishing caffeinated and decaffeinated products.

Another crucial issue is the degree of homogeneity in the customer population. We would like to consider customers identical in the sense that one set of utility weights applies to all. This will not imply that all customers behave in the same manner because some of the attribute values will vary from customer to customer.

For example, households are well known to be heterogeneous with respect to purchase probabilities because of differing brand preferences or loyalties. To model this we can introduce a customer loyalty attribute defined as a weighted sequence of past purchases of a brand. To hold down the number of constants, we can make loyalty a common attribute across brands, i.e., use the same  $b_j$  coefficient for all.

Customers also frequently show a preference or loyalty to a particular package (cans versus bottles in soft drinks) or size (one pound versus two pounds or three pounds in coffee). Such preferences can be introduced as customer attributes in the same way as brand loyalty. By bringing in past customer purchase behavior as an explanatory variable through loyalty attributes, we conveniently model purchase probability heterogeneity while treating customers homogeneously.

#### 4. Regular Ground Coffee

Coffee is a frequently purchased product actively marketed by manufacturers and retailer alike and so makes an excellent subject for building and testing a choice model. Price changes are relatively common because of fluctuations in commodity markets. Coffee is also a traditional supermarket loss leader with frequent promotions during which the store takes one or more coordinated marketing actions such as reducing price, putting up a special display or running an advertisement in the local media.

##### 4.1. Data

Our database consists of ground coffee store and panel records from four Kansas City supermarkets for the 78-week period September 14, 1978 to March 12, 1980. The data have been collected by Selling Areas-Marketing, Inc. (SAMI) and kindly made available for this research.

The store sales data contain weekly item movement for each UPC in the category as well as the shelf price for each item each week. The customer panel has about 2,000 households, each of which has indicated it makes 90% or more of its purchases at one of the stores collecting the data. A single purchase record contains the household number, the date of purchase, the UPC, and the price paid. The panel has been cleaned to eliminate households with reporting gaps that indicate lack of store loyalty and also to omit households joining the panel in the middle of the period. The resulting static sample has been further reduced to exclude light and nonusers of ground coffee, these being defined as households that made less than five purchases during the year and a half of data. Quite a few families remain, however, and, of them, 200 have been chosen at random. A group of 100 forms the calibration sample and the second 100 a hold-out sample, reserved for testing the final model. The calibration group made 1037 purchases in the chosen brand-sizes during the 32-week period, March 8, 1979 to October 17, 1979 used for calibration. Sixteen of the purchases had missing store data leaving 1021 usable purchases. The holdout group made 921 purchases in the corresponding 32-week period plus another 666 in a 20-week post period October 18,

1979 to March 5, 1980 used for projection. Five of these had incomplete data leaving 1582 purchases for evaluative purposes.

In addition to the information generated by the store computers, we have local newspaper advertisements that assist in identifying promotional activity.

##### 4.2. Alternatives

Our selection of product alternatives draws on the research of Urban et al. (1984) about the structure of the coffee market. They divide the coffee market into six product groups defined by ground/instant, caffeinated/decaffeinated and, within instant, freeze dried/nonfreeze dried. To obtain a relatively homogeneous set of alternatives we restrict consideration to regular (caffeinated) ground coffee.

Another issue is brand versus size. In the Kansas City market, the popular sizes are one pound and three pound which we shall call “small” and “large.” (Included in “small” are certain 13 and 14 ounce sizes advertised as producing the same number of cups as a pound.) Preliminary examination of switching behavior yielded no evidence to suggest that customer choice was hierarchical on either brand or size. However, different sizes of the same brand are clearly different products from both retailer’s and customer’s point of view. Customers show distinct size loyalty and retailers promote sizes separately. Therefore we model brand-sizes.

Ground coffees come as regular, drip, and automatic. However, given the brand-size, all grinds are priced the same and promoted together. A household’s choice depends primarily on its coffee making equipment. We therefore combine UPC’s across grinds to construct the brand-size alternatives.

Two low share brand-sizes (less than 1% of the purchases of the calibration sample) were dropped for lack of observations, leaving as the set of alternatives the eight largest selling brand-sizes.

Sizes		Brands				
Small	A	B	C	D	E	
Large	A	B	C			

##### 4.3. Attribute Variables

The latent customer utility will be expressed as a linear function of attribute variables consisting of one unique attribute for each alternative and a set of attributes common across all alternatives. Each attribute variable will have a value for each alternative for each purchase, for example, the  $n$ th purchase of customer  $i$ .

(1) *Unique to alternative.* The utility function for a brand-size will include an additive constant specific

to that alternative. This is accomplished by a set of dummy variables:

$$x_{0k}^i(n) = \begin{cases} 1 & \text{for alternative } k, \\ 0 & \text{otherwise.} \end{cases}$$

Each alternative,  $k$ , has its own variable except that one of them, say the  $N$ th, must be omitted to avoid a singularity in the maximum likelihood estimation, since, as previously remarked, the utilities are undetermined to the extent of an additive constant. The value of  $b_{0N}$  can be taken as zero. The resulting brand-size specific constants,  $b_{0k}$ , serve to capture any uniqueness an alternative has that is not captured by other explanatory variables insofar as describing average choice behavior over all observations is concerned. One consequence of introducing the  $b_{0k}$  is that the average predicted probability of an alternative will equal its share of purchases over the observations. No other alternative-specific variables will be used.

(2) *Common across alternatives.* The major control variables we shall consider for the retailer are regular (depromoted) price, the presence or absence of a promotion for the brand-size and, if there is a promotion, the amount of price-cut (possibly zero) during the promotion. The first of these attribute variables is the regular price in the absence of promotion:

$$x_{1k}^i(n) = \text{regular (depromoted) price of brand-size } k \text{ at time of customer } i\text{'s } n\text{th coffee purchase (dol/oz).}$$

If brand-size  $k$  is on promotion at the purchase occasion, its regular or depromoted price is its price prior to the start of the promotion. The second variable is promotion:

$$x_{2k}^i(n) = \begin{cases} 1 & \text{if brand-size } k \text{ was on promotion at time of customer } i\text{'s } n\text{th coffee purchase,} \\ 0 & \text{otherwise.} \end{cases}$$

Unfortunately, no data on feature or display activity are available for the stores in our sample. We do, however, have information on three events that would be expected to accompany a promotion: (1) short-term (one to three weeks) price reduction, (2) mention of the item in the store's weekly newspaper advertisement, and (3) an unusually high short-term movement for the product, chosen as store sales greater than two standard deviations above the average for the product without promotion. Since any one of these events alone may be a faulty indicator, we infer a promotion only if two out of three of these events take place in the given week. Although the unusual movement indicator seems potentially circular as a purchase explainer, we note that (1) unusual movement is

determined from store data not the panel data being explained, (2) the movements are very large, typically 3 to 10 times normal sales, and are clearly the result of store phenomena, not chance, and (3) unusual movement alone never defines a promotion because of the 2 out of 3 rule. Note also that any difficulty in identifying promotions here is a data problem not a model problem. That is, field observation can determine whether a product is featured and the amount of its display even though this information is not in our present set. Thus, even though the lack of in-store observation may conceivably have introduced some bias in the present situation, the model formulation seems adequate.

The 0–1 promotion variable will miss some of the variation among promotions, such as the difference in response between an end-aisle display and a small shelf-talker. However, we can explain some of the variation by adding a depth of discount or promotional price cut variable to the model:

$$x_{3k}^i(n) = \text{promotional price cut on brand-size } k \text{ at time of customer } i\text{'s } n\text{th coffee purchase (dol/oz).}$$

The variable is zero when there is no promotion of  $k$ . Thus the actual price of brand-size  $k$  is always  $x_{1k} + x_{3k}$ .

An important marketing question is whether a promotional purchase is, in some sense, as good as an ordinary purchase. For example, will a household that switches to a brand when it is on promotion be as likely to repurchase the brand at a later time as would be the case for a nonpromotional purchase? Shoemaker and Shoaf (1977), Dodson et al. (1978), and Jones and Zufryden (1981) have found that a promotional purchase decreases the likelihood of a subsequent purchase of that brand. We can look for this effect by introducing lagged promotion variables.

$$x_{4k}^i(n) = \begin{cases} 1 & \text{if customer } i\text{'s previous purchase of coffee was a promotional purchase of an alternative with the same brand as brand-size } k, \\ 0 & \text{otherwise;} \end{cases}$$

$$x_{5k}^i(n) = \begin{cases} 1 & \text{if customer } i\text{'s second previous purchase of coffee was a promotional purchase of an alternative with the same brand as brand-size } k, \\ 0 & \text{otherwise.} \end{cases}$$

A further set of variables depends on characteristics of the customer. As observed earlier, much of the heterogeneity in purchase probability over a population, be it called preference, habit or loyalty, can be captured by observing past behavior. We choose the



word loyalty to describe the customer's tendency to repurchase the same brand size. Let

$x_{6k}^i(n)$  = brand loyalty for brand of brand-size  $k$  for  $n$ th coffee purchase of customer  $i$ .

We define

$$x_{6k}^i(n) = \alpha_b x_{6k}^i(n-1) + (1 - \alpha_b) \begin{cases} 1 & \text{if customer } i \text{ bought} \\ & \text{brand of alternative } k \text{ at} \\ & \text{purchase occasion } (n-1) \end{cases} \begin{cases} \\ \\ 0 \end{cases} \text{ otherwise}$$

Operationally, therefore, loyalty is taken to be the exponentially weighted average of past purchases of the brand, treated as 0–1 variables. The carryover constant is  $\alpha_b$ . To start up brand loyalty, we set  $x_{6k}^i(1)$  to be  $\alpha_b$  if the brand of alternative  $k$  was the first purchase in the data history of customer  $i$ , otherwise  $(1 - \alpha_b)/(\text{number of brands} - 1)$ , thus insuring that the sum of loyalties across brands always equals 1 for a customer.

Size loyalty is analogous:

$$x_{7k}^i(n) = \text{size loyalty for the size of brand-size } k \text{ for } n\text{th coffee purchase of customer } i. \\ = \alpha_s x_{7k}^i(n-1) + (1 - \alpha_s) \begin{cases} 1 & \text{if customer } i \text{ bought size} \\ & \text{of alternative } k \text{ at coffee} \\ & \text{purchase } (n-1) \end{cases} \begin{cases} \\ \\ 0 \end{cases} \text{ otherwise}$$

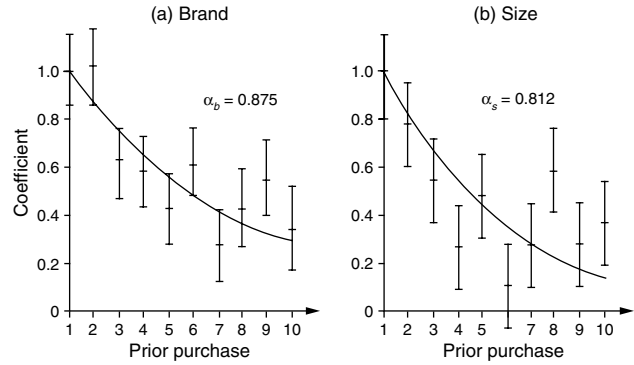
where  $\alpha_s$  is the carry-over constant for size. Initialization is analogous. The sum of loyalties over sizes is also 1.

For all variables except the alternative-specific dummies, the coefficients,  $b_j$ , are the same for all brand-sizes. Thus in calculating customer  $i$ 's utility the same coefficients will be used for every brand-size with respect to attributes such as price, promotion, loyalty, etc. This gives the model remarkably few parameters, considering the number of products addressed. Using the same coefficients across brand-size does not imply the same response to control variables. Response will depend on the whole marketing environment and on the customers' loyalty variables.

#### 4.4. Calibration

The calibration database contains 100 households with 1021 coffee purchases over 32 weeks spread across eight brand-sizes. In addition, 718 purchases over the previous 25 weeks have been used to initialize the loyalty variables. Each purchase will be treated as an observation so that we are combining cross-section and time-series data. This makes the loyalty variables particularly important since they

**Figure 3** The Smoothing Parameters in the Loyalty Functions Are Fit to Coefficients Generated by Lagged Purchase Variables



Note. Bars indicate  $\pm$  one standard error.

carry not only much of the cross-sectional heterogeneity but also a good part of the purchase-to-purchase dynamics.

An immediate question, however, is how to pick the smoothing constants  $\alpha_b$  and  $\alpha_s$  for constructing the loyalty variables. The answer is: to pick trial values then refine them. Approximate  $\alpha$ s ( $\alpha_b = \alpha_s = 0.75$ ) are used and the model is developed with all other major explanatory attributes. Following this, the brand loyalty variable is replaced by ten dummy variables, each denoting whether or not the brand was purchased on the  $n$ th prior occasion ( $n = 1, \dots, 10$ ). The relative coefficient sizes indicate the impact of the  $n$ th prior purchase. The results, normalized to make the first coefficient unity, appear in Figure 3(a). The solid line shows the exponential decay selected. The value of  $\alpha_b$  has been chosen to make the differences between the line and the coefficients average to zero. The same process is repeated with lagged size dummies, as shown in Figure 3(b). The exponential function is quite satisfactory since the number of points more than a standard deviation from the line is close to what would be expected (5 actual versus 6 expected). The corresponding  $\alpha_b$  and  $\alpha_s$  define the loyalty variables. By varying the  $\alpha$ s over a grid of values and recalculating the maximum likelihood estimates of the other parameters until the likelihood function is also maximized by the  $\alpha$ s we could have all parameters as maximum likelihood estimates. We elected not to go to these rather substantial further efforts.

As may be seen the decay rates for brand and size loyalty are comparable. The brand coefficients decay slightly more slowly (higher  $\alpha$ ) suggesting somewhat more loyalty to brand than size, a conclusion to be supported again later.

#### 5. Calibration Results

The calibrated logit model fits the coffee data well and provides information on a number of marketing issues. Table 1 shows the coefficients.

**Table 1** Calibration of Coffee Model for 6 Specifications with Increasing Number of Variables Plus a Simulated Aggregate Model

Specification	S1	S2	S3	S4	S5	S6	S7
$U^2$	0	0.24	0.46	0.47	0.47	0.48	0.22
Brand loyalty		2.78 (22.1)	3.47 (21.9)	3.47 (21.5)	3.79 (21.3)	3.92 (21.6)	
Size loyalty		2.12 (14.6)	2.74 (16.1)	2.72 (15.9)	2.74 (15.9)	2.97 (15.9)	
Promotion			2.22 (15.5)	2.00 (13.6)	2.07 (13.9)	2.11 (14.1)	1.40 (11.5)
Promotional price cut			18.12 (5.5)	29.66 (7.2)	29.20 (7.1)	29.21 (7.1)	26.98 (8.0)
Regular (depromoted) price				−26.36 (−6.0)	−26.49 (−5.9)	−29.94 (−6.6)	−28.02 (−7.8)
Price promotional purchase					−0.60 (−3.5)	−0.22 (−1.3)	0.62 (4.5)
Second prior promotional purchase					−0.72 (−3.9)	−0.46 (−2.47)	0.49 (3.3)
Brand Size Constants							
Small A	0.28 (2.2)	−0.12 (−0.9)	−0.06 (−0.4)	0.03 (0.2)	0.01 (0.1)	−0.09 (−0.6)	0.48 (3.5)
Large A	0	0	0	0	0	0	0
Small B	0.48 (3.9)	0.28 (2.0)	0.11 (0.7)	0.19 (1.2)	0.20 (1.2)	0.10 (0.6)	0.41 (3.1)
Large B	−0.06 (−0.4)	0.24 (1.6)	0.05 (0.3)	−0.06 (−0.4)	−0.05 (−0.3)	−0.08 (−0.5)	−0.35 (−2.3)
Small C	0.99 (8.8)	0.14 (1.0)	0.31 (2.0)	0.57 (3.6)	0.62 (3.8)	0.41 (2.5)	1.49 (11.6)
Large C	0.38 (3.1)	0.01 (0.1)	0.03 (0.2)	0.22 (1.3)	0.27 (1.6)	0.12 (0.7)	0.59 (4.2)
Small D	−1.25 (−6.1)	−1.23 (−5.3)	−0.66 (−2.6)	0.15 (0.5)	0.10 (0.3)	−0.05 (−0.2)	0.10 (0.4)
Small E	−1.83 (−6.8)	−1.52 (−5.2)	−0.76 (−2.3)	−1.5 (−4.1)	−1.47 (−4.1)	−1.72 (−4.7)	−1.88 (−6.3)
Log likelihood ( $N = 1,021$ )	−1,896	−1,440	−1,025	−1,007	−1,002	−977	−1,488

Note. Table entry shows attribute coefficients with  $t$ -statistic beneath in parentheses.

### 5.1. Quality of Fit

To indicate the relative contribution of various attributes and to investigate the stability of the coefficients against changes in model specifications, the final model has been built up a few variables at a time. We can follow the changes in  $U^2$ , the amount of uncertainty explained by the model.

The specification S1 contains only the brand-size dummy variables. The effect of this is to make each household's purchase probability for a brand-size the same and equal to that brand-size's share of total purchase. Since this is our chosen null model,  $U^2 = 0$ . In S2 the addition of the brand and size loyalty variables produces a large jump to  $U^2 = 0.24$ , demonstrating, as expected, that the loyalty variables explain much purchase behavior across households. Notice the high  $t$ -statistics. Specification S3 introduces the promotion variables and another big jump to  $U^2 = 0.46$ . Bringing in regular price in S3 increases  $U^2$  only a little to 0.47. Clearly regular price is not moving share around the way promotion does. However, a chi-squared test of

S4 relative to S3 shows statistical significance at the 0.01 level. When the prior promotional purchase variables are added in S5,  $U^2$  increases slightly but after rounding is still 0.47.

The last step is to revise  $\alpha_b$  and  $\alpha_s$  as described earlier and rerun with the new loyalty variables. The result is S6, which we take as the final specification. It has  $U^2 = 0.48$  indicating improved loyalty functions as a result of the iteration. Although the prior promotional purchase variable is not significant in a statistical sense, we retain it because the second prior is. (We have checked the third prior and it is not.)

Notice the growth of the brand loyalty coefficient as more uncertainty is explained. This happens because the utility scaling increases to hold the residual variance of  $\epsilon_k$  constant in (1), as discussed in §2.1. If enough variance were explained, the coefficients would become very large and the predicted probabilities would approach 0 and 1.

The brand-size constants would be zero if we had found variables that explained all the differences

among brands and sizes. Obviously this would be difficult to achieve and we have not, but notice that, in going from  $S1$  to  $S6$ , most of the brand-size constants and their  $t$ -values decline.

Except for the brand-size constants, the coefficients, once introduced, tend to be rather stable throughout the various specifications, a healthy sign indicating that collinearity does not seem to be a serious issue. The only exception is the increase in the promotional price cut coefficient when regular price is introduced. The increase seems reasonable, however, since the regular price is the reference point from which the price cut is measured and its absence might reasonably hinder the explanatory ability of the cut.

An important practical question is whether a household-level model really offers any substantial advantages over a more aggregate model. Specification  $S7$  investigates this issue by dropping the principal customer-specific variables of  $S6$ , namely, the brand and size loyalties. All other explanatory variables are the same. We see a dramatic drop in  $U^2$  to 0.22. This clearly demonstrates the power of the customer-level loyalty variables for explaining purchase behavior across households. The coefficients for price, promotion, and price-cut stay about as before. This is probably comforting, although, if they were different, the disaggregate variables explain more uncertainty and would be preferred. The prior promotion coefficients change to a positive sign and become statistically significant. This is not too surprising because these variables are now the closest thing in the model to the loyalty variables. We also note that the brand-size constants increase. This is presumably because they carry the burden of fixing average share, some of which role was previously picked up by loyalty. In summary, the quality of the model deteriorates and the coefficients become more difficult to interpret, thereby providing solid support for a disaggregate model.

Pictures illustrating quality of fit are always helpful. A direct comparison of the probability of purchase calculated by the model and the actual purchase outcomes is of little value since the latter are either 0 or 1. However, the expected number of purchases of a brand size in a time period is easily calculated from the model since the category purchases are given. The expected number divided by total purchases gives the expected share of purchases which can then be compared to actual share. Figure 4 does this, showing predicted versus actual purchase shares for the eight brand-sizes over the calibration period. The curves are gratifyingly close.

As discussed earlier, the multinomial logit model implicitly assumes “independence from irrelevant alternatives” (IIA). McFadden et al. (1977) have devised a residuals test to evaluate whether the IIA

assumption holds in a given case. The underlying idea is that violation of the assumption will cause systematic errors in predicted choice probabilities. The observations are divided into cells in a systematic way and a goodness of fit test applied to each brand-size. In Appendix 1 we present the results of applying the procedures to our model. None of the eight individual tests shows a systematic error at the 0.05 level of significance. We conclude that violations of the IIA assumption are not a serious problem. We feel, however, that a more important test of the model will be its performance on a holdout sample of customers.

## 5.2. Discussion of Coefficients

All the coefficients have the algebraic signs that would be expected.

The relative importance of attributes in explaining market behavior is of interest. The coefficient magnitudes per se are not too instructive because their units vary. Two better indicators are the contribution to  $U^2$  and the  $t$ -statistic. Given a fixed sample size, the relative magnitudes of the  $t$ -statistics offer an indication of explanatory importance. In part this is because the numerator of  $t$  is the coefficient itself and so increases with its size. In addition, the denominator is the coefficient's standard error, which will tend to decrease if the data for the attribute have large variance, other things being equal. Such variance also opens the possibility of explaining considerable behavior.

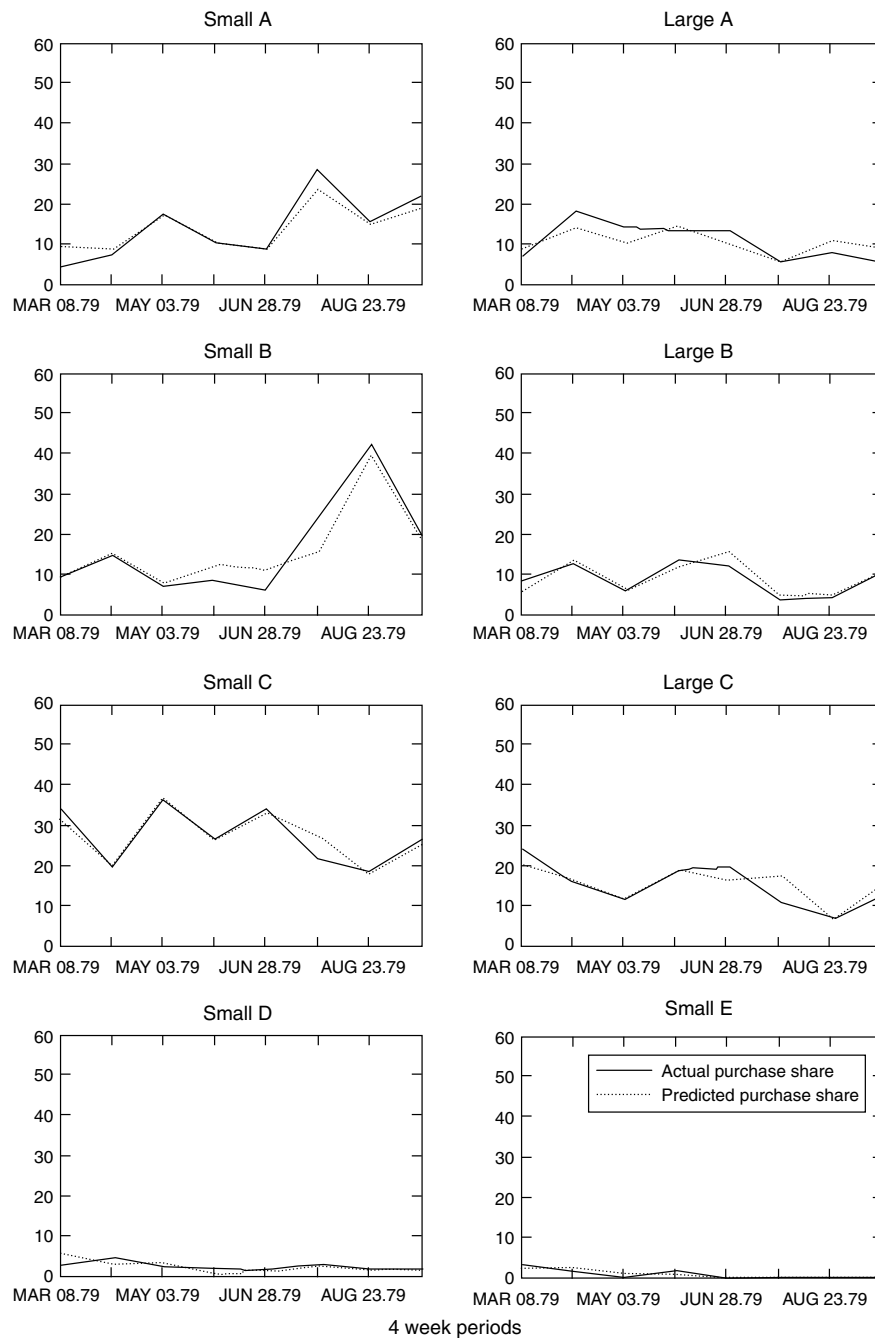
Using these indicators we find brand and size loyalty most important. Brand has a larger coefficient and  $t$ -value than size, as well as a larger carry-over constant as noted earlier. However, size loyalty is extremely important, probably much more than manufacturers and retailers realize.

The next most important attribute is the 0–1 promotion variable, indicating that the attention that manufacturers and retailers give to promotion is not misplaced. After promotion a drop takes place to the two price variables, regular price and promotional price cut, and finally another drop to the prior promotional purchase.

The brand-size constants form a special group. They can represent unique product qualities and/or specification errors. As already mentioned, if the other explanatory variables are doing a nearly perfect job, these constants should be close to zero. Large  $A$  has been taken as the reference point and is zero by definition. Most of the others are small. However, it is revealing to compare brand-size constants with shares, arranged in decreasing order:

	Small C	Small B	Large C	Small A	Large A	Large B	Small D	Small E
Share of purchases:	28.4	17.1	15.5	14.0	10.6	9.8	3.0	1.6
Brand-size constant:	0.41	0.10	0.12	−0.09	0	−0.08	−0.05	−1.72

**Figure 4** Predicted Share of Purchases Tracks Actual Share Closely for the Calibration Sample over the Calibration Period



Perhaps a modest relationship exists, although only Small C and Small E at the extremes have significant  $t$ -values (2.5 and  $-4.7$ ). One would conjecture that an explanatory variable may be missing. A top candidate would be quality of display during a promotion. Small E stands out even further than this would be expected to explain and, in fact, is a slightly different product with certain unique qualities.

Looking at the control variables further, we see that regular and promotional price coefficients have about the same magnitude. We do not expect this a priori

because the promotional price is likely to be advertised in the newspapers and by signs within the store. Similarity of coefficients does not necessarily imply similarity of price response, however, because of the other terms in the utility function.

Promotion is the most interesting control variable. Not only is its effect large but usually it is accompanied by a price cut that enhances total impact. Some marketers believe the effect of promotions to be entirely due to price but that view is not supported by our results. Even with a price cut of zero, a large

probability gain is calculated. This is consistent with much experience that special displays (e.g., end-aisle or free standing positions) play a major role in customer response.

The past promotion variable is directed at an important question in package goods marketing: Is a customer who buys a product on promotion less likely to repurchase that product than a customer who buys it under ordinary circumstances? The answer is needed for assessing the true value of promotional activity. The work of Shoemaker and Shoaf (1977), Dodson et al. (1978), and Jones and Zufryden (1981) indicates that the promotional purchase has a negative effect on subsequent purchase probability. The present model permits an explicit assessment of the effect. A promotional purchase affects subsequent buying through brand loyalty, size loyalty and the prior promotional purchase variables. Suppose that a customer buys a particular brand-size on promotion. Then the contribution of that purchase to the next period's utility as a result of the loyalty variables is (using coefficients from S6 of Table 1, the  $\alpha$ s of Figure 3, and the definitions of  $x_6$  and  $x_7$ ):

$$(3.92)(0.125)(1) + (2.97)(0.188)(1) = 1.05.$$

However, this contribution is diminished by the prior promotional purchase variable:

$$-(0.22)(1) = -0.22.$$

This leaves a net plus of 0.83 relative to a situation of not having purchased the brand-size at all. Carrying the process to the next purchase, the contribution of the original purchase to utility through the loyalty variables is:

$$(3.92)(0.875)(0.125)(1) + (2.97)(0.812)(0.188)(1) = 0.88$$

but is diminished by the second prior promotional purchase variable

$$-(0.46)(1) = -0.46$$

for a net of 0.42. The negative effect of the prior promotional purchase variables is in agreement with previous authors, but net contribution over the two periods remains positive. In other words the calculation says that the customer is more likely to purchase the promoted brand-size than if it had not been purchased in the first place, but less likely than if it had been a nonpromotional purchase.

## 6. Testing

The holdout sample of 100 households (1609 purchases) provides an opportunity to challenge the calibrated model. Tracking share of purchases by 4-week

period will be our primary method of evaluation. Not only does share loom important in practice but it also varies dramatically over time and over brand-size. We ask several questions:

(1) How well does the model predict brand-size share in the hold-out sample during the same time period used in the calibration?

(2) How well does the model predict share during time periods subsequent to the calibration?

(3) Can the model, using constants derived from data that mixes together all stores, predict shares within individual stores? In comparing predicted versus actual, we need a measure of random variation, especially for low share brand-sizes, whose number of purchases may be very few in a particular four weeks. Accordingly, we calculate a standard error of predicted share. For observation  $i$  and any given brand-size, the model predicts a probability of purchase  $p_i$ . Given the null hypothesis that the calibrated model is correct, the actual purchase is binomially distributed. Letting  $s$  denote the predicted share and  $n$  the number of observations,

$$s = \sum_{i=1}^n p_i / n,$$

$$SE(s) = \left[ \sum_{i=1}^n p_i(1 - p_i) \right]^{1/2} / n,$$

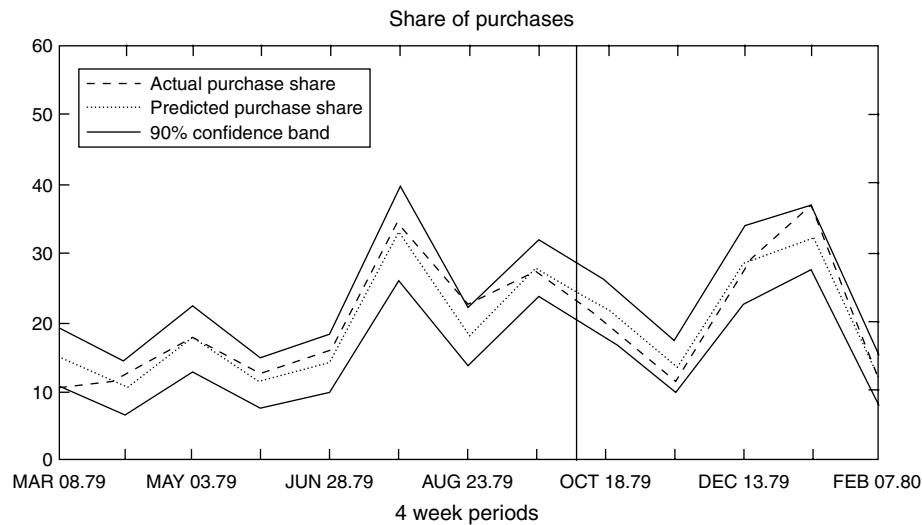
where  $SE(s)$  is the standard error of share.

Figure 5 displays an example of hold-out tracking, the case of Small A. The dotted line in the middle is the model prediction, the upper and lower solid lines are  $\pm 1.64$  SE and so form approximately a 90% confidence band. The loyalty variables are initialized in a pre-period September 14, 1978 through March 7, 1979. Tracking then takes place during the period from March 8, 1979 through October 17, 1979 previously used by the calibration sample and on into a new, forecast period October 18, 1979 through March 5, 1980.

As may be seen, agreement is remarkably good. A general upward trend is captured. So is a promotional peak and then a downward trend that is interrupted temporarily by another promotion. The model is impressive in tracking these ups and downs. We credit this to detailed information about the store environment at the time of purchase (namely, price, promotion, and promotional price cut for *all* competitors) plus information about the customer's purchase history summarized in the brand and size loyalty variables.

Of special interest is the good quality of tracking in the period October 18 to March 5, since here we not only are using a hold-out sample but also are forecasting outside the time period used by the calibration sample. In the forecast period we continue to employ

**Figure 5** Tracking of Small A Purchase Share in a Holdout Sample of 100 Customers Shows Actual Share Falling Almost Entirely Within the Confidence Band



*Notes.* The vertical line separates the calibration time period from the forecast period. During the forecast period loyalty variables are constructed from simulated purchases.

the actual prices and promotions of all the brand sizes (i.e., the model is not attempting to predict what marketing decisions will be made). A dilemma arises, however, over the loyalty variables. Loyalty uses purchases in its construction and yet the purpose of the holdout sample is to predict purchases. We can, of course, simply use the purchases just for the construction of the loyalty variables. This seems justified since we are always predicting ahead in time relative to the purchases entering the calculation. We use this procedure for March 8 to October 17, 1979. However, for October 18, 1979 to March 5, 1980, we adopt a new, Monte Carlo method so as to be above reproach on the issue. We construct a synthetic purchase history for each customer as follows: When an actual purchase is needed to update the loyalty variables, we draw a random number and use it to select a brand-size to be purchased, this being done so that the probability of picking a brand-size conforms to its model-calculated purchase probability. As the forecast period proceeds, the synthetic purchase history unfolds for each customer and enters the loyalty variables.

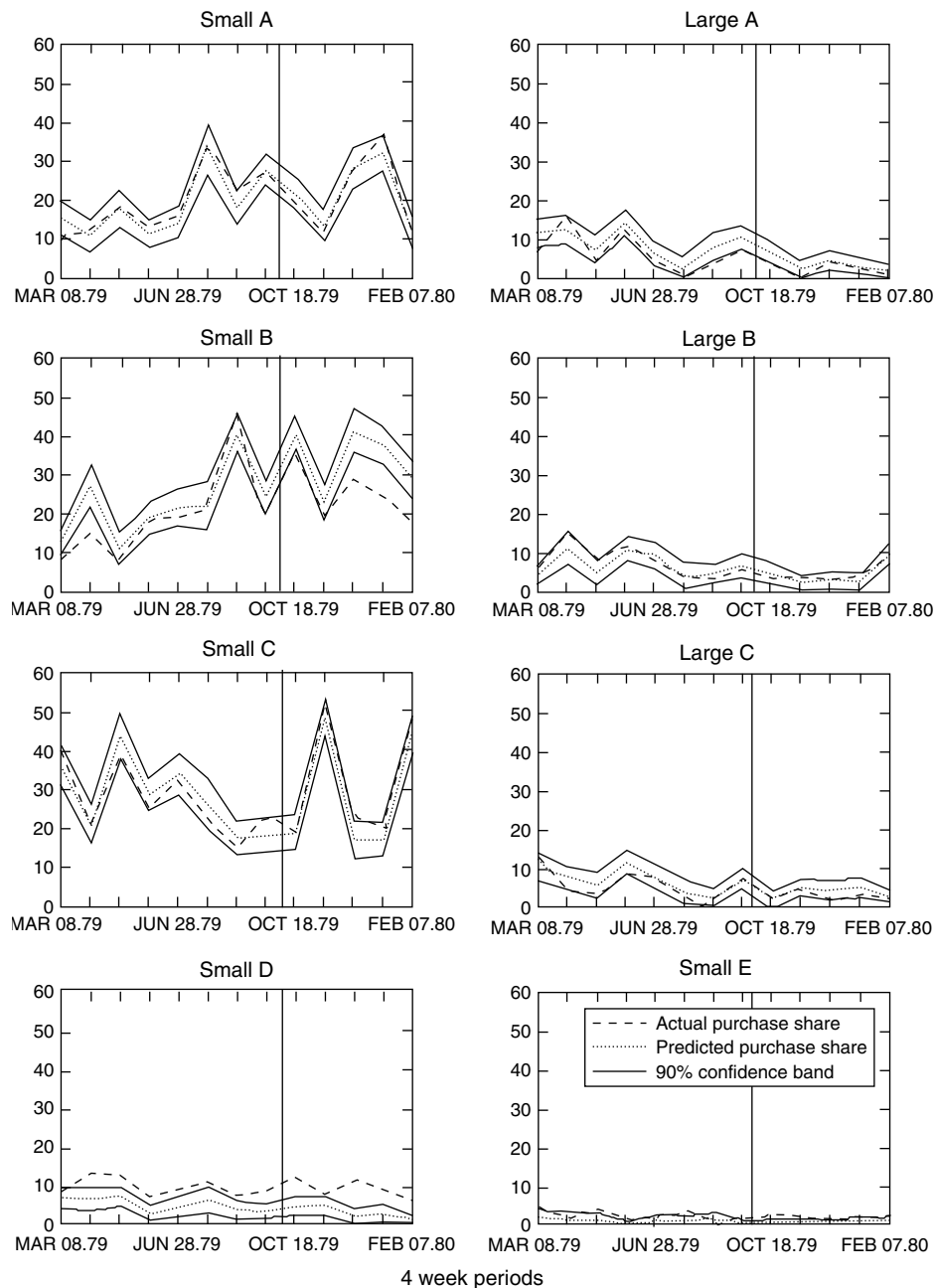
We see in Figure 5 that, even with the severe test of a new set of customers and a conservative loyalty calculation, tracking is good in the post-calibration periods. Figure 6 displays corresponding plots for all brand-sizes. The model follows the turns and trends of share quite well. Enough points lie outside the 90% confidence band, however, to suggest the model has not captured all phenomena and, indeed, we know some variables are missing.

In most cases the share *changes* are followed within a standard error or so but sometimes the absolute

level is missed. The deviation of Small B in the forecast period is an example. It would appear from the quality of fit in general that the model captures quite well the average effect of, say, promotion, but an individual promotion may be more or less successful than average. The one for Small B in the middle of the forecast period appears to have been less successful. The deviation persists after the promotion because the model increases the customers' loyalties on the basis of their assumed promotional purchases, whereas the actual customers experience no such effect. Despite the level difference, however, the directional changes continue to be correct.

Small D shows a level difference that puts the actual share entirely above the confidence band. Presumably, this is due to a miscalibration of the brand-size constant. On checking back, it turns out that the calibration sample contains only 31 purchases of Small D out of a total of 1021 during the calibration period (3.0% share). Despite random selection of customers for the two samples, the holdout group has many more purchases of Small D, about 93 out of 916 (10.2% share) in the same period. A larger brand-size constant would produce a better representation of the underlying population that both samples are drawn from and would certainly improve the appearance of the Small D plot. Be that as it may, the share changes are tracked well so that, with respect to the normative questions of what happens when the decision variables are changed, the Small D plot is reassuring.

Another opportunity to test the model arises from the presence of different stores in the sample. The panel population consists of store-loyal customers from four stores, each of which has its own individual personality. One of them is a large conventional

**Figure 6** Tracking of Purchase Share in the Holdout Sample Is Quite Good for Almost All Brand-Sizes

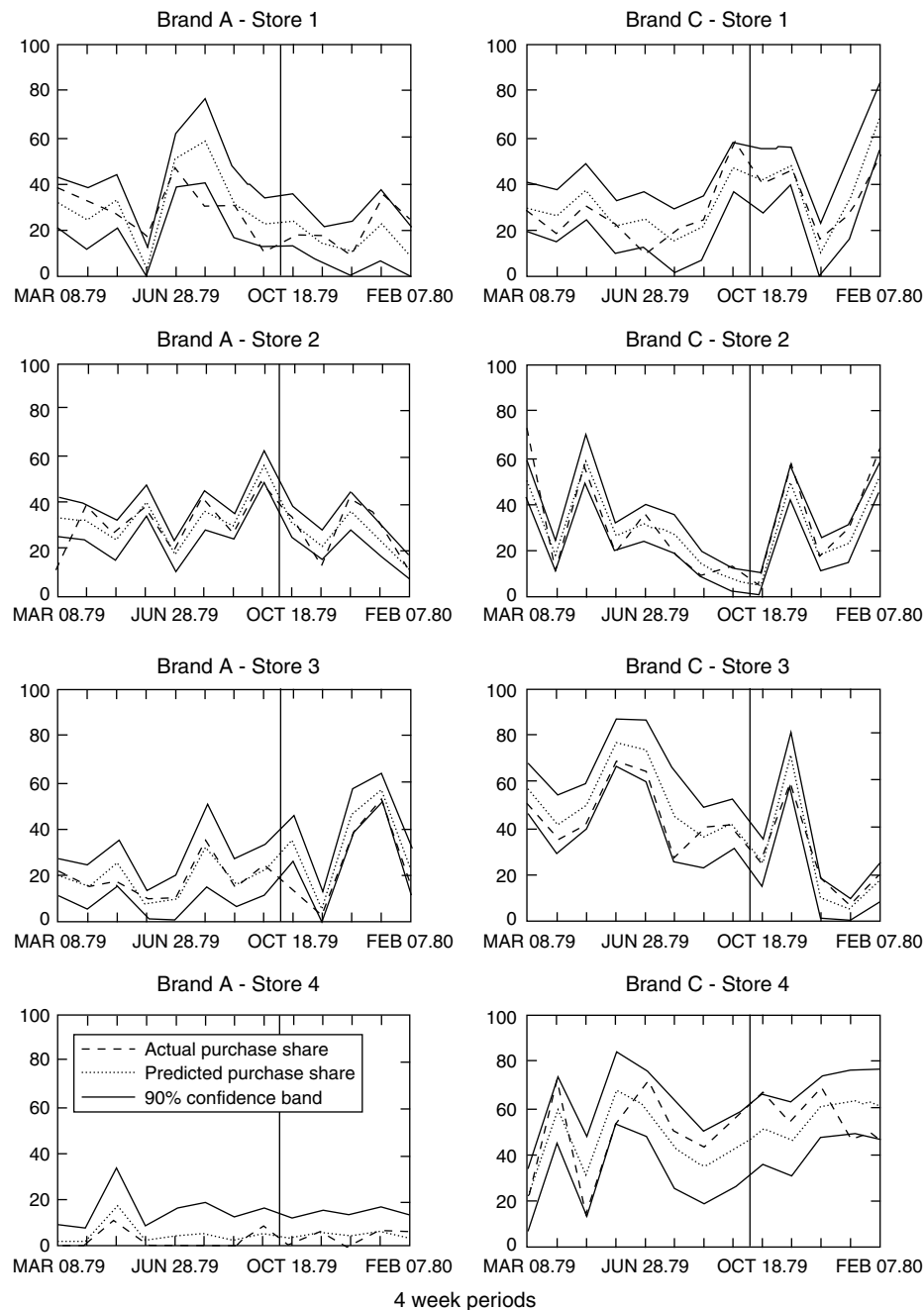
supermarket; another a smaller, more neighborhood market; still another is a warehouse store; and the last is a conventional store with unconventional merchandising practices. The calibration process takes no account of the differences among stores beyond whatever manifestations these may generate through values of the attribute variables, nor does the calibration consider that different types of households may shop at different stores.

Breaking the customers down by store reduces sample sizes. To compensate we aggregate over sizes to brand level and also focus on two high share

brands A and C. Figure 7 presents tracking for these brands in the four stores. The format is as in Figure 5: holdout sample, predicted and actual shares with confidence band. The results are very satisfactory, with the model predicting behavior well across the four stores. Notice in particular the tracking of the decline of Brand C in Store 3. The implication is that if we know what is happening to the control variables in the store, we can predict share quite well.

The parsimony of the model deserves emphasis. Omitting the brand-size constants which, as we have discussed, determine only average share, the coeffi-

**Figure 7** Tracking of Purchase Shares of Brands A and B in Four Different Supermarkets Is Quite Good Despite Different Store Characteristics



cients of the model are only seven in number and are exactly the same for every customer, store, and brand-size. Nevertheless, they combine with the control and environmental variables to produce widely different patterns of share that correspond well to actual behavior in the hold-out sample.

## 7. Market Response to Control Variables

Practical use of the model requires measures of how the market responds to the retailers' actions.

### 7.1. Calculating Response

We have a calibrated model at the individual customer level but aggregate market response is of more interest to a decision maker and requires further computational effort. First of all, even though the logit coefficients are the same for all customers, individual responses will differ greatly depending on the store environment and prior loyalties. Table 2 illustrates this by showing that, in a hypothetical two-product market, a customer with equal loyalties to both products displays a price elasticity of share equal to  $-2.2$ , whereas a customer with  $0.8$  loyalty to one of



**Table 2** Although the Model Coefficients Are the Same for All Customers, Response Can Be Quite Different Depending on Attribute Values at Time of Purchase

Hypothetical two product market: A and B	
Loyalty coefficient: 7.0	
Price coefficient: −30 oz/dol	
No other attributes.	
Price response of customer with equal loyalties	
Loyalty to A = loyalty to B = 0.5	
Price of A = price B = 0.15 dol/oz	
Probability of choosing A = $e^{(3.5-4.5)} / (e^{-1} + e^{-1}) = 0.5$	
10% price reduction of A to 0.135 dol/oz	
Probability of choosing A = 0.61	
Price elasticity for A = −2.2	
Price response of customer with unequal loyalties	
Loyalty to A = 0.8, loyalty to B = 0.2	
Price of A = price of B = 0.15 dol/oz	
Probability of choosing A = 0.985	
10% price reduction of A to 0.135 dol/oz	
Probability of choosing A = 0.991	
Price elasticity for A = −0.06	

the products has an elasticity of −0.06 to that product. This is an example of the S-shape of Figure 2 in action. Analogously, although the control variables in our eight-product market have the same coefficients for all brand-sizes, the market response for a particular product will depend on the actions of the other brand-sizes and the distribution of loyalties across the customers.

Analytically, determination of aggregate market response calls for an integration of the customer response function over a joint distribution of customer loyalties, prices, and promotions for every brand. One standard approach to such a high dimensional integration is Monte Carlo sampling. A simpler and easier method to implement is to use the actual customers, time periods and attribute variables of the data. By changing, say, the regular price of a brand-size by 1% over the whole time period and observing the corresponding change in share, we obtain an aggregate share response to regular price.

Many different response calculations are possible. Some examples are (1) regular price elasticity of share, (2) the cross elasticity of one brand-size's share to another's price, (3) share response to promotion, (4) the cannibalization of one size of a brand by the promotion of another, and (5) the elasticity of price during a promotion.

Each of these can be evaluated as a short-term effect, i.e., as an alteration of choice probabilities at the purchase occasion on which the variable is changed, or as a long-term effect, taking also into account the future impact of changed loyalty. We restrict ourselves here to three examples of short-term response: regular price elasticity, promotion response, and promotional

**Table 3** Short-Term Market Response to Three Different Control Variables

	Purchase share (%)	Regular (depromoted) price elasticity of share	Share increase from promotion with median price cut (%)	Promotional price cut elasticity of share
Small C	28.4	−1.9	173	−1.1
Small B	17.1	−2.3	334	−1.7
Large C	15.5	−2.3	273	−1.7
Small A	14.0	−2.5	362	−1.8
Large A	10.6	−2.7	363	−1.8
Large B	9.8	−2.6	502	−2.0
Small D	3.0	−3.4	409	−3.2
Small E	1.6	−2.4	568	−2.4

price cut elasticity. The results will illustrate a number of important points. The calculations are made as follows:

(1) *Regular Price Elasticity.* Using the calibration sample of the 32-week calibration period, all the regular prices of, say, Small A are increased by 1% and other variables held constant. The probability of choosing Small A is calculated for all purchases. This produces a new purchase share for Small A over the calibration period. The percent change in purchase share divided by the present change in price is taken as the short-term regular price elasticity.

(2) *Response to a Promotion with a Median Price Cut.* Consider again Small A. For every purchase occasion calculate the probability of purchase without a promotion. Calculate the average share for Small A. Repeat the process but compute the probability of purchase when Small A has a promotion with a median price cut. All other attributes are held constant throughout. Our measure of short-term promotion response is the percent increase of the second purchase share over the first.

(3) *Promotional Price Cut Elasticity.* For every purchase occasion calculate probability of purchase with a Small A promotion having a median price cut with 1% of regular price added back in. Calculate Small A purchase share. Take the percentage difference from the share with promotion and median price cut and divide by 1%. The result is the short-term promotional price cut elasticity.

The results are in Table 3.

## 7.2. Discussion

A first observation is that response varies quite widely by brand-size. This was expected even though all brand-sizes use the same coefficients for the control variables, because the market response calculations take into account the complete environment, including competitive conditions, brand loyalties, etc. The variation in response by brand-size has significant implications for retailers and manufacturers.

Differences in price elasticity, for example, suggest different pricing policies across brand-sizes.

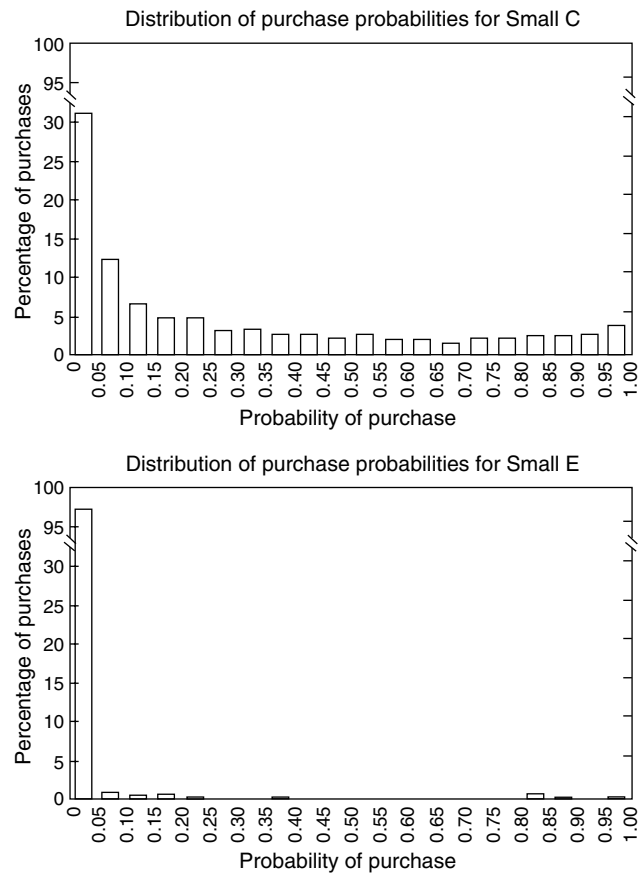
The next point is that the response measures show a definite pattern. Table 3 has been arranged so that the brand-sizes are listed in order of decreasing share of purchases. A glance down the table shows that the magnitudes of the response numbers tend to increase with decreasing share. For example, in each case the average of the four lower brand-sizes is larger in absolute magnitude than that of the higher four. The relationship is not perfect, thereby happily leaving room for brand individuality. Nevertheless, as a percentage of their own share, high share brand-sizes are less sensitive to their own control variables than low share brand-sizes. On the other hand, we can multiply the three response measures of Table 3 by the corresponding shares to obtain absolute share increments for the changes in control variables. If this is done, it will be discovered that the high-share brands tend to obtain a larger share increment than the low-share brands for the same marketing action.

We argue that these relationships between sensitivity and share are forced by the structure of the logit model. Details appear in Appendix 2. We find that the patterns of Table 3 exist at the individual purchase level and so their manifestation in aggregate is not surprising. We also find that the magnitude of regular price elasticity will be largest when all individual purchase probabilities are identical and equal to purchase share. Heterogeneity of purchase probabilities then lowers the magnitude of elasticity.

The nature of the heterogeneities in practice is illustrated in Figure 8, which shows histograms of purchase probabilities across purchases for the highest and lowest share brand-sizes. As may be seen, each product has a group of high probability purchasers, a broad range of intermediate values indicating switchers, and a cluster of near-zero probabilities. The latter is very large for the low share brand-size. A high purchase probability for a product implies a high utility in the logit model, often as a result of high loyalty. As mentioned in the discussion of the S-shaped curve of Figure 3, customers with high or low utilities for a given brand-size will have corresponding purchase probabilities that are relatively insensitive to marketing actions, thus reducing the overall responsiveness of the product.

In summary, the structure of the logit model tends to produce short-term sensitivities to marketing variables, which, when expressed in elasticity form, decrease in absolute value with increasing purchase share. A broad distribution of loyalties (and therefore probabilities of purchase) reduces sensitivity relative to a set of identical customers. Neither of these statements is dependent on the coffee market or the specific brands involved, but instead is only a function of the model form.

**Figure 8** The Distribution of Purchase Probabilities Across Purchases for the Highest and Lowest Share Brand-Sizes Shows a Slight Bimodality with Higher Extremes and a Broad Middle



As a final observation about market response we remark on the completeness of the model. It gives information not just about a single brand-size but about the whole market. We could supplement Table 3 with cross-elasticities and other response measures as described earlier. The completeness permits the study of strategic issues, for example, the consideration by a company not only of its own actions but also of various possible competitive reactions.

## 8. Conclusions

The multinomial logit model has provided an excellent representation of the regular ground coffee market at the individual customer and retail store level. Coefficients of the model are statistically significant, many of them highly so. The calibrated model predicts the behavior of a hold-out sample of customers satisfyingly well. A remarkable feature of the model is its parsimony. The major coefficients, namely, those for brand and size loyalty and the control variables, are the same across all brand-sizes and all customers, and yet the model ably predicts brand-sizes with widely different shares, follows different trends and

turns over time and tracks brand performance in different types of stores. The results are not perfect and, reasonably so, since various marketing phenomena are missing (for example, display quality, couponing, and media advertising) as we know from other studies that these actions influence purchases. Nevertheless, the results here seem very promising.

The success of the modeling effort appears to be due to the micro detail and competitive completeness of the scanner panel data. People, not markets, respond to the actions of the retailers and manufacturers. The greater variability of the explanatory variables at the individual level offers richer opportunities for calibrating response than the corresponding store or market data.

Manipulation of the calibrated model yields share response to several marketing variables. Here we have only scratched the surface. However, it is clear that, because we are modeling the actions of all brand-sizes, the answers produced by the model will depend strongly on the questions it is asked. In particular the answers will depend on the customer loyalties and competitive actions assumed.

Much work remains to be done. A major missing feature is the modeling of the purchase occasion itself. Our work has focused entirely on share, whereas certain market actions, notably promotion, tend to shift purchases in time and therefore at least temporarily expand the market.

The model needs other extensions. Presently, it describes what happens inside the retail store. This is not adequate for the retailer, who is relatively unconcerned about competition among brands, except perhaps for house brands, but is very concerned about competition from other retailers. (See Little and Shapiro 1980.) Correspondingly, although manufacturers are strongly interested in interbrand competition, the control variables of our model are those of the retailer and only indirectly influenced by the manufacturer. Various steps should be taken to extend the results closer to the decision makers' needs.

A good representation of the coffee market is encouraging, but coffee is just one of the many product categories. We do not yet know whether the strong assumptions of the multinomial logit will hold up in more complicated situations where, for example, product differentiation is greater or variety seeking is common or the customer maintains a portfolio of products. We expect that new modeling issues will emerge and require ingenuity to resolve.

Nevertheless, in the coffee case we see the suggestion of a new understanding of the interplay between brand franchise and marketing actions. Here is a model, frugal in parameters, that fits the data well and reveals a distinct pattern between share of purchases and market response. Since neither the relationship nor tracking is perfect, there appear to be

idiosyncratic effectiveness variables by brand and purchase occasion and, of course, other marketing variables await inclusion in the model. Nevertheless at the level of general understanding of these markets, the picture emerges that a well-entrenched brand-size has a set of loyal customers who make its share relatively insensitive to certain marketing actions, at least in the short run. At the same time the market also contains a pool of switchers less loyal to any brand-size who respond fairly readily to changes in marketing variables. Reliable knowledge about the interplay among attributes and their ramifications for brand strategy now seem to be within striking distance of our emerging marketing technology.

### Acknowledgments

The authors thank Selling Areas Marketing, Inc. (SAMI) for providing the scanner panel data on which this work is based. The authors are especially appreciative of the help of Kenneth Silvers, Frank Smith, Valerie Gager, and Jack Moffly.

### Appendix 1. Test for Independence from Irrelevant Alternatives

The multinomial logit model implicitly assumes "independence from irrelevant alternatives" (IIA). McFadden et al. (1977) present a residuals test that we shall apply to evaluate whether the assumption reasonably holds in our set of data. The motivating idea of the test is that violation of the IIA property will cause systematic errors in the choice probabilities. The procedure first calculates the probabilities by the calibrated model for, say, alternative  $j$  for each of the 1021 observations. The probabilities are then ranked and sorted into some number of cells, roughly the same number of observations in each cell. For each cell we calculate an expected number of choices of  $j$  from the probabilities and compare it with the actual number in a goodness of fit test. The statistic

$$\chi^2 = \sum_{m=1}^M (S_m - N_m \bar{P}_{jm})^2 / N_m \bar{P}_j \quad \text{where}$$

$m$  = index of cell,

$M$  = total number of cells,

$S_m$  = number of actual choices of  $j$  in cell,

$N_m$  = total number of observations in cell,

$\bar{P}_{jm}$  = average probability for alternative  $j$  in cell  $m$ , and

$\bar{P}_j$  = average probability for alternative  $j$  in total sample, has an asymptomatic distribution bounded by  $\chi^2$  distributions with  $M - 1$  and  $M - K - 1$  degrees of freedom where  $K$  is the number of estimated parameters. The test statistics are not independent across alternatives.

The test was run for each brand-size, dividing the observations into 50 cells each time or about 20 observations per

cell. The resulting  $\chi^2$  were

	$\chi^2$		$\chi^2$
Small A	24.6	Small C	17.9
Large A	18.5	Large C	18.6
Small B	33.0	Small D	9.5
Large B	17.0	Small E	30.2

The degrees of freedom for the upper bound  $\chi^2$  statistic is 49 with critical (0.05) level 66.1 and for the lower bound 35 and 49.7. As may be seen the  $\chi^2$  for each brand-size falls well below both critical levels. Therefore no departure from IIA is detected.

## Appendix 2. Effect of Share on Price Response

Table 3 shows that the magnitude of short-term market response, when expressed as a percentage of purchase share, increases with decreasing share. After multiplying by share, response decreases with share. We claim that these phenomena are the result of the structure of the multinomial logit model. Our analysis will be done for the case of regular price.

Consider first a single customer and purchase decision. The expected share of purchases for one occasion is simply the probability of purchase. Let

$$\begin{aligned} \eta_{kk}^i(n) &= \text{the regular price elasticity of purchase share} \\ &\quad \text{for brand-size } k \text{ on purchase occasion } n \text{ of} \\ &\quad \text{customer } i. \\ &= \frac{x_k^i(n)}{p_k^i(n)} \frac{\partial p_k^i(n)}{\partial x_k^i(n)} \end{aligned} \quad (A1)$$

where  $x_k^i(n)$  is the regular price and  $p_k^i(n)$  the purchase probability of brand-size  $k$  on purchase occasion  $n$  of customer  $i$ .

Purchase probabilities are calculated from the multinomial logit

$$p_k^i(n) = e^{v_k^i(n)} / \sum_j e^{v_j^i(n)} \quad (A2)$$

with

$$\begin{aligned} v_k^i(n) &= a_k^i(n) + bx_k^i(n)y_k \\ &= \text{utility of brand-size } k \text{ for customer } i \text{ on } n\text{th purchase occasion,} \\ a_k^i(n) &= \text{effects of all other constants and variables on utility} \\ &\quad \text{for the same purchase occasion,} \\ b &= \text{coefficient of regular price,} \\ y_k &= \text{a parameter introduced in order to change all prices} \\ &\quad \text{of brand-size simultaneously and proportionally.} \\ &\quad \text{Nominal value} = 1.0. \end{aligned}$$

For immediate purpose  $y_k = 1$ . Substituting the derivative of (A1) into (A2) and simplifying yields

$$\eta_{kk}^i(n) = bx_k^i(n)[1 - p_k^i(n)]. \quad (A3)$$

Therefore, on an individual basis, the absolute magnitude of regular price elasticity increases as share decreases. This explains the aggregate effect found in Table 3. Further  $|p_k^i(n)\eta_{kk}^i(n)|$ , the incremental share for a percentage change in price, decreases with decreasing share on an individual

basis over the range of  $0 \leq p_k^i(n) \leq 1/2$ . Provided that the preponderance of the probabilities falls in this range, as will be the case for brand-sizes having shares less than 1/2, the second type of pattern in Table 3 can be expected.

Although (A3) explains the patterns found, further insight can be gained about aggregate elasticity. Let

$$\begin{aligned} m_k &= \text{expected purchase share of brand-size } k \\ &= \frac{1}{N} \sum_i \sum_n p_k^i(n) \end{aligned} \quad (A4)$$

where  $N_i$  = the number of purchases by customer  $i$  and  $N = \sum N_i$ . Further let

$$\begin{aligned} \eta_{kj} &= \text{cross-elasticity of purchase share of } k \\ &\quad \text{with respect to price of } j \\ &= \frac{y_j}{m_k} \frac{\partial m_k}{\partial y_j}. \end{aligned} \quad (A5)$$

From (A2)

$$\frac{\partial p_k^i(n)}{\partial y_j} = bx_j^i(n) \begin{cases} p_k^i(n) - p_k^i(n)^2, & j = k, \\ -p_k^i(n)p_j^i(n), & j \neq k. \end{cases} \quad (A6)$$

Using (A5), (A4), and (A6), and setting  $y_j = 1$ , we obtain

$$\eta_{kk} = \frac{b}{m_k} \frac{1}{N} \sum_i \sum_n x_k^i(n)[p_k^i(n) - p_k^i(n)^2], \quad j = k, \quad (A7)$$

$$\eta_{kj} = \frac{b}{m_k} \frac{1}{N} \sum_i \sum_n x_j^i(n)p_k^i(n)p_j^i(n), \quad j \neq k. \quad (A8)$$

Although (A7) looks a little complicated, we can use it to learn the effect of heterogeneity of  $p_k^i(n)$  on  $\eta_{kk}$ . To do this, we first ask how we would pick the distribution of  $p_k^i(n)$  to maximize  $|\eta_{kk}|$  subject to the constraint that purchase share is fixed, i.e.,  $(1/N) \sum \sum p_k^i(n) = m_k$ . (For simplicity, we study the case  $x_k^i(n) = x_k$ , a constant price for all purchases.) Mathematically, the problem is equivalent to finding  $\{u_i\}$  to maximize  $\sum (u_i - u_i^2)$  subject to  $\sum u_i = Nm_k$ ,  $0 \leq u_i \leq 1$ ,  $i = 1, \dots, N$ . In turn, this is equivalent to minimizing  $\sum u_i^2$  subject to the same constraints. It can be shown, for example by dynamic programming, that the solution is  $u_i = m_k$ .

This means that the magnitude of regular price elasticity  $|\eta_{kk}|$  will be largest for a set of customer purchases which have identical purchase probabilities all equal to the aggregate purchase share of the group. Heterogeneity reduces  $|\eta_{kk}|$  from there. In fact, pushing the  $p_k^i(n)$  to all 0s and 1s would give  $\eta_{kk} = 0$ . For the identical probability case, (A7) reduces to

$$\eta_{kk} = bx_k(1 - m_k).$$

Using the calibration value  $b = -29.9$  and the average coffee price in the market, 0.165 dollars/ounce, we obtain  $\eta_{kk} = -4.93(1 - m_k)$ . This shows  $|\eta_{kk}|$  to be increasing with decreasing share, as expected. Comparison of values from this formula with those in Table 3 shows the effect of heterogeneity on  $|\eta_{kk}|$ . As an example, Small C would have  $|\eta_{kk}| = 3.5$  by this formula but is 1.9 by Table 3.

To summarize, the logit structure inherently produces the patterns of response obtained. Further, the greater the dispersion of the individual purchase probabilities away from  $p_k^i(n) = m_k$  toward all 0s and 1s, the lower the magnitude of elasticity.

## References

- Ben-Akiva, M. 1973. Program for maximum likelihood estimation of the multinomial logit model. Transportation Systems Division, Department of Civil Engineering, MIT (January).
- Dodson, J. A., A. M. Tybout, B. Sternthal. 1978. Impact of deals and deal retraction on brand switching. *J. Marketing Res.* 15(February) 72–81.
- Domencich, T. A., D. McFadden. 1975. *Urban Travel Demand*. North-Holland, Amsterdam.
- Gensch, D. H., W. W. Recker. 1979. The multinomial, multiattribute logit choice model. *J. Marketing Res.* 16(February) 124–132.
- Hauser, J. R. 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: An information theoretic approach. *Oper. Res.* 26(May) 406–421.
- Hlavac, T. E., Jr., J. D. C. Little. 1966. A geographic model of an Urban automobile market. D. B. Hertz, J. Melese, eds. *Proc. Fourth Internat. Conf. Oper. Res.*, Wiley-Interscience, New York, 302–311.
- Jones, J. M., F. S. Zufryden. 1980. Adding explanatory variables to a consumer purchase behavior model: An exploratory study. *J. Marketing Res.* 17(August) 323–334.
- Jones, J. M., F. S. Zufryden. 1981. Relating deal purchases and consumer characteristics to repeat purchase probability. *J. Market Res. Soc.* 23 84–99.
- Little, J. D. C., J. F. Shapiro. 1980. A theory for pricing nonfeatured products in supermarkets. *J. Bus.* 53(July, Part 2) S199–S209.
- Luce, R. D. 1959. *Individual Choice Behavior*. John Wiley and Sons, New York.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.
- McFadden, D. 1981. Econometric models of probabilistic choice. C. F. Manski, D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA, 198–272.
- McFadden, D., K. Train, W. B. Tye. 1977. An application of diagnostic tests for the independence of irrelevant alternatives property of the multinomial logit model. *Transportation Res. Record* 637 39–45.
- Punj, G. N., R. Staelin. 1978. The choice process for graduate business schools. *J. Marketing Res.* 15(November) 588–598.
- Shoemaker, R. W., F. R. Shoaf. 1977. Repeat rates of deal purchases. *J. Advertising Res.* 17(April) 47–53.
- Silk, A. J., G. L. Urban. 1978. A pre-test-market evaluation of new packaged goods: A model and measurement methodology. *J. Marketing Res.* 15(May) 171–191.
- Theil, H. 1969. A multinomial extension of the linear logit model. *Internat. Econom. Rev.* 10(October) 251–259.
- Urban, G. L., P. L. Johnson, J. R. Hauser. 1984. Testing competitive market structures. *Marketing Sci.* 3(2) 83–112. [Originally cited as: Urban, G. L., P. L. Johnson, R. H. Brudnick, J. R. Hauser. 1982. Modeling of competitive market structure: A methodology based on forced switching probabilities. Sloan School Working Paper 1236-81, MIT (November).]
- Yellott, J. I., Jr. 1977. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment and the double exponential distribution. *J. Math. Psych.* 15 109–144.