WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

WHEN AND WHAT TO BUY:
A NESTED LOGIT MODEL OF COFFEE PURCHASE

Peter M. Guadagni
Information Resources, Inc.
Four Embarcadero Center, San Francisco, CA 94111

John D.C. Little
M.I.T. Sloan School of Management
Cambridge, MA 02139

August 1987 WP 1919-87

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

WHEN AND WHAT TO BUY:
A NESTED LOGIT MODEL OF COFFEE PURCHASE

Peter M. Guadagni
Information Resources, Inc.
Four Embarcadero Center, San Francisco, CA 94111

John D.C. Little
M.I.T. Sloan School of Management
Cambridge, MA 02139

August 1987 WP 1919-87

ABSTRACT

On a shopping trip to a supermarket a customer may purchase a product in a given category and, if so, buys a particular brand. A previous paper by the authors models the brand choice part of this process. A multinomial logit model describes the selection of brand and size given that the customer makes a category purchase. Explanatory variables include store actions, such as price and promotion, and customer characteristics, such as brand and size loyalty. We now extend the formulation to include the decision to make a purchase in the category on a shopping trip. This additional step not only provides a more complete description of the buying process but also makes possible a better calculation of sales response by including the effect of marketing actions on category sales as well as brand share.

The methodology employed is the generalization of the multinomial logit known as the nested logit. The shopper's decision includes two components: the selection of the category and the choice of the brand-size combination. The model of brand-size choice is essentially that of our earlier paper. The category choice introduces new variables including household inventory, category price, and the attractiveness of purchasing a product now as opposed to later. Calibration of the nested logit is done by sequential estimation. The model, applied to regular ground coffee data from a UPC scanner panel, tracks sales well in a holdout sample, both at the aggregate and individual levels. Use of the model to calculate short term market response to promotion demonstrates the phenomenon that brand sales can increase because of the expansion of the category as well as share.

1. INTRODUCTION

In an earlier paper the authors have modeled a customer's choice of product within a category as affected by retail store actions and prior purchasing behavior (Guadagni and Little, 1983). Principal retailer control variables are price, presence or absence of store promotion, and amount of promotional price cut. Customer behavioral variables are prior loyalty to brand and size. The basic methodology is the multinomial logit, with calibration and testing done on UPC scanner panel data for the regular ground coffee category.

In the earlier paper, we take as known that a customer makes a purchase in the category. This restriction is now removed by extending the model to include the decision to buy a product in the category on a particular shopping trip. Whereas the earlier model was driven by purchases, each representing an opportunity to choose a particular brand and size, the new model is driven by shopping trips, each representing an opportunity to buy within the category. Since we are dealing with products that are bought repeatedly, the decision is really whether to buy now or later. The theory used for the extended model is the nested logit as described by Ben-Akiva and Lerman (1985). Calibration is by sequential estimation. The present paper expands and deepens the work initially reported by Guadagni (1983).

The importance of modeling the category purchase is two-fold: First, it enhances our understanding of the variables that affect customer decisions. Second, it will permit a more complete calculation of sales response to marketing actions, since marketing often affects category sales as well as brand shares. Figure 1, which shows the coffee sales that we plan to model, demonstrates that much often goes on at the total category level.

For motivation and exposition we view purchase as a two stage process: first the category is chosen, then the particular product. We note, however, that the theory does not require this actually to be the process. The model considers the customer's ultimate alternatives (the purchase of any of the products on the shelf or the postponement of purchase) in a way that acknowledges the different interactions among the products as a group and between the set of products and postponed purchase.

The category purchase decision has its own set of explanatory variables. These include measures of household inventory, category price, and the attractiveness of buying on that shopping trip as affected by the product choice variables at the time. In addition, for our coffee application, we model a major environmental disturbance that occurred during the time period of our data, namely, a freeze in the Brazilian coffee producing regions.

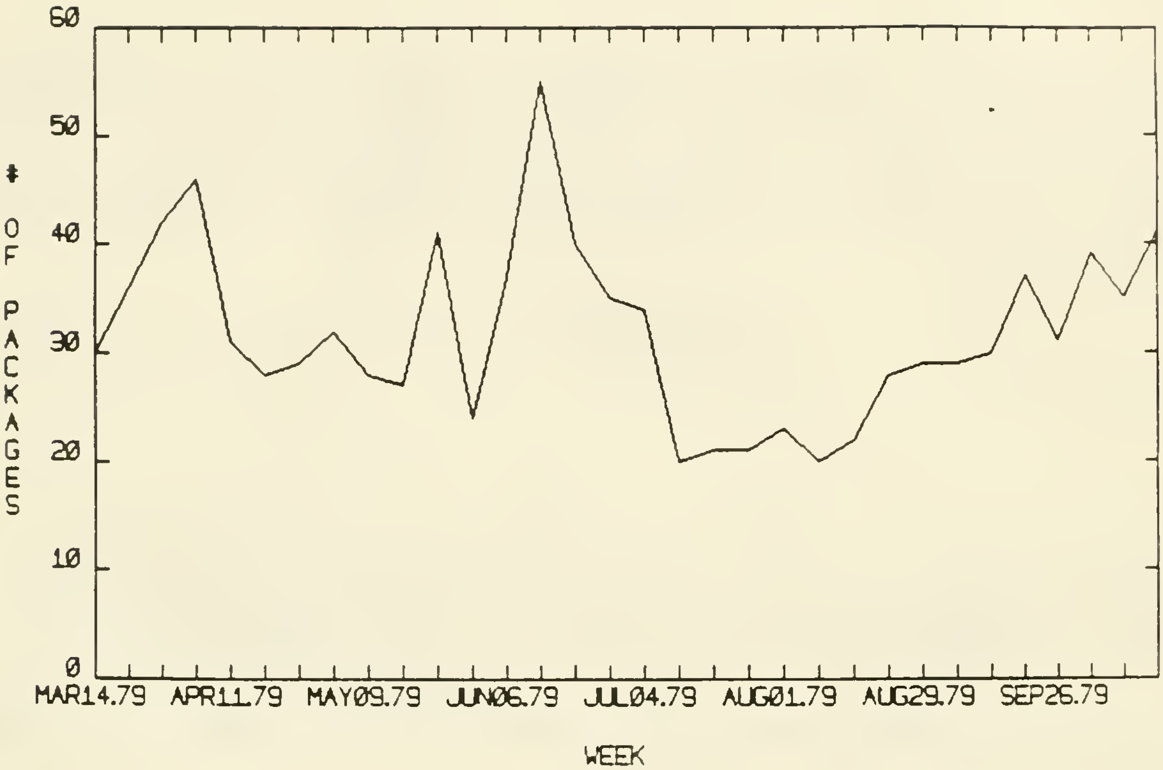


Figure 1. The number of regular ground coffee purchases each week in a panel of 100 households shows substantial change over a 32 week period.

2. THE NESTED LOGIT

Figure 2 displays a decision tree for a customer on a shopping trip. The customer may be viewed as deciding sequentially when to buy and then what to buy but with interaction between the decisions. The top level of the tree represents the choice of whether to buy within the category now or later. The second level is the choice of product given the category purchase. We have dotted the subtree under buy-later to indicate that it is not known in advance exactly what the products and their attributes will be, a point we shall return to later.

Letting a_1 and a_2 represent the category choices buy-now and buy-later and $b_1, b_2, \ldots, b_B$ the products available, the customer's choice set is

$$\{(a_1,b_1),(a_1,b_2),\ldots\ldots(a_2,b_1),\ldots(a_2,b_B)\} \qquad (1)$$

These alternatives may be viewed as a multidimensional choice set with dimensions a and b. Among the ways to model such a set, the nested logit seems especially appropriate. We shall describe the underlying assumptions of the model and several major properties. For a detailed discussion and relevant derivations, see Ben Akiva and Lerman (1985), Chapter 10.

In general the utility of customer i for the choice (a,b) is taken to be (suppressing the customer index i)

$$u_{ab} = v_a + v_b + v_{ab} + e_a + e_b + e_{ab} \qquad (2)$$

where v_a, v_b, and v_{ab} = the systematic components of utility related to category, product, and the category-product combination, respectively;

e_a, e_b, and e_{ab} = unobserved components of utility for category, product, and category-product combination.

For a nested logit model containing the category choice decision at the top level as in Figure 2, we shall assume that $var(e_b)$ is negligible compared to $var(e_a)$. Consequently, e_b is dropped from (2). Sufficient conditions to obtain the standard nested logit choice probabilities are then:

1. e_a and e_{ab} are independent for all a and b in the customer's choice set.

2. The terms e_{ab} are independent and identically distributed with a double-exponential (Gumbel) distribution that has a scale parameter μ_b.

3. e_a is distributed so that $\max_b u_{ab}$ is double-exponentially distributed with scale parameter μ_a.

Figure 2. The customer's decision tree on a shopping trip. The customer may be viewed as deciding to buy in the category now or later and, if now, choosing a product within the category.

The value of μ_b can be set to unity as a normalization and we shall henceforth do so.

As shown in Ben Akiva and Lerman (1985), the probability of buying product b given that the category choice is a is

$$p(b|a) = \{\exp(v_{ab} + v_b)\}/\sum_j \exp(v_{aj} + v_j) \quad (3)$$

and the probability of making the category decision a is

$$p(a) = \{\exp[(v_a + v'_a)\mu_a]\}/\sum_j \exp[(v_j + v'_j)\mu_a] \quad (4)$$

where

$$v'_a = \ln \sum_b \exp(v_b + v_{ab}) \quad (5)$$

Thus product choice probability is an ordinary multinomial logit model involving only parameters from the second level of the tree. Category choice would also be an ordinary logit for its level except for the introduction of v'_a and μ_a.

The term v'_a is extremely interesting. Mathematically, it is the natural log of the denominator of the product choice logit expression. It can be shown (Ben Akiva and Lerman, 1985, Chap. 10) that, except for an additive constant, it is the expected value of $\max_b (v_b + v_{ab} + e_a + e_{ab})$. Or it can be described as the systematic component of the maximum utility of the subset of product alternatives that involve a.

As may be seen from (4) and (5), v'_a takes information from the product choice level and inserts it into the category choice decision. In fact, the larger the utilities v_b and v_{ab} at the product level, the larger v'_a. The larger v'_a, the more likely is alternative a to be the category choice selected.

Various descriptive names have been given to the quantity v'_a. McFadden (1981), for example, calls it the "inclusive value". Because of its property that, as it increases, so does the probability of choosing that value of a, we shall call it <u>category attractiveness</u> in our application.

An important goal is to express v_a, v_b, and v_{ab} in terms of explanatory attributes, especially marketing variables. We do this linearly. Taking into account some of the specialization of our application, it suffices to consider

$$v_a = \sum_j d_j z_{ja} \quad (7)$$

$$v_b = \sum_j b_j x_{jb} \quad (8)$$

where the z_{ja} and x_{jb} are explanatory variables for the category and product level choices, respectively, and the d_j and b_j are coefficients that will be estimated from data. In our application v_{ab} will not be needed.

Estimation of the parameters of the nested logit can be done sequentially as follows:

Step 1. Apply standard multinomial logit estimation to the conditional choice model $p(b|a)$ at the bottom level. This determines $\{b_j\}$.

Step 2. Calculate the category attractiveness:

$$v'_a = \ln \sum_b \exp[\sum_j b_j x_{jb}] \tag{9}$$

for each value of a.

Step 3. Using v'_a as a separate independent variable, estimate, via standard logit programs, the logit parameters μ_a (the coefficient of v'_a) and $\{\mu_a d'_j\}$ for the category choice model $p(a)$, where $\{d'_j\}$ are the coefficients in a linear expression containing the explanatory variables.

From Step 3 we see that we can think of v'_a as just another explanatory variable, assign it a d_j to represent μ_a, set $d_j = \mu_a d'_j$ for the other variables, and then work with (7). We note that, despite the notation, μ_a does not depend on a, but rather is a generic constant for the a level.

3. CALIBRATION AND TESTING

As just discussed, the hierarchical structure of the nested logit permits sequential estimation of its parameters. Calibration in nested fashion is statistically consistent and quite convenient, since logit estimation programs are common. The method is not as statistically efficient as full information maximum likelihood procedures, but the latter would require extensive new program development. McFadden (1981) provides a transportation example in which both methods are used. Although results are generally similar, the full information case provides some improvement in fit and produces certain shifts in parameter values. We shall stick to the nested logit in our work but the full information calculation is worthy of investigation.

Measures of calibration quality will help evaluate the model. Logit packages generate asymptotically valid t-values for individual parameter estimates. An overall measure of model fit is U^2 (equivalent to ρ^2). This quantity provides a measure of uncertainy explained by the model and is given by

$$U^2 = 1 - L/L_0 \tag{10}$$

where L is the log-likelihood of the test model and L_0 the log-likelihood of a reference model. The value of U^2 varies between 0 and 1. If the test model does no better than the reference model, $U^2 = 0$. If the test model is a perfect predictor of choice, $U^2 = 1$. The reference model to be used here is that the probability of choosing an alternative equals its aggregate share of choices.

Another measure of quality is predicted vs. actual. Aggregate plots are helpful, comparing market share over time with average predicted probability. An individual-based measure is $\bar{P}$, the expected value of correct predictions. This is calculated by weighting the actual choice (taken as unity) with the probability of making that choice and averaging over all choice occasions. Perfect prediction would score as 1, total failure as 0.

We shall also use a holdout sample for evaluation. Model assumptions are many and calibration procedures complex so that a good test is often some form of predicted vs. actual comparison on a holdout sample. Parameters estimated from one group of customers are used in the model to predict the behavior of another group.

4. SCANNER PANEL DATA FOR COFFEE

Regular ground coffee makes a good category for applying the model. Coffee receives many retail promotions, is subject to price changes, and is frequently purchased. The data for this study is the same as that used in our earlier paper and was collected by SAMI (Selling Areas Marketing Inc.) from four Kansas City supermarkets. SAMI has very kindly made the data available for academic research.

Included in the database are store and panel information for a 74 week period from September 14, 1978 to February 12, 1980. The store data contains weekly store sales and shelf price for each item in the regular ground coffee category. The panel data contains the items purchased, the date of the purchase, and the price paid for each panelist. The panel data also indicates the date of each shopping trip made by the household. From the total Kansas City panel, a static sample has been created by removing households who joined in the middle or who had extensive reporting gaps. From the static sample we have drawn 200 coffee-using households at random. Coffee-using has been defined as making five or more coffee purchases in the time period. The model is calibrated from the shopping and purchasing behavior of 100 households and the second 100 is kept as a holdout sample for testing and evaluation.

In addition to the information gathered by the store computers, we have local newspaper advertisements that help identify store promotional activity.

5. PRODUCT CHOICE WITHIN CATEGORY: WHAT PRODUCT TO BUY

The bottom level of the tree represents the choice of a product given that the customer has decided to buy coffee. The product choice model presented here, except for minor simplications, is the same as that of Guadagni and Little (1983). A brief description of the model and a complete discussion of the changes follow. For more detail see the original article.

5.1 Product Choice Model Specification

Observations. Each purchase of a can of coffee represents an observation in the product choice model. The variable representing the choice records for each individual is:

$$y^i_k(n) = \begin{cases} 1 & \text{if customer i chooses alternative k on the nth choice occasion.} \\ 0 & \text{otherwise.} \end{cases}$$

Alternatives. The set of alternatives contains the eight largest selling brand-sizes available in Kansas City during the time period of the analysis. Included are all brand-sizes having one percent or more of all purchases. Table 1 lists the products and their shares of purchases within the calibration sample.

	Share of Purchases (%)	
Size:	Small	Large
Brand:		
Maxwell House (A)	14.0	10.6
Butternut (B)	17.1	9.8
Folgers (C)	28.4	15.5
Folgers Flaked (D)	3.0	
Mellow Roast (E)	1.6	

Table 1. Brands and sizes used in the product choice model with their shares of purchases in the calibration sample. Letters in parentheses indicate designation used in Guadagni and Little (1983). Small refers to one pound and Large to three pound except for Folgers Flaked which contains 13 ounces and Mellow Roast which contains 12.

Notice that, although Folgers Flaked has the same brand name as Folgers, it is being treated as a separate brand. Folgers Flaked is produced by a different technology, has different physical properties (it is more concentrated) and is promoted separately from the parent brand. One would therefore expect it to behave as a separate product.

Product attribute variables. The utility of each brand-size alternative is expressed as a linear function of attribute variables. These variables are separated into those unique to the alternative and those common across all alternatives.

(1) Unique to an alternative.

The utility function for a brand-size alternative includes an additive constant specific to that alternative. This is accomplished by a set of dummy variables, one for each brand-size, K (K=1,2,...,B), except that one of them must be omitted to permit estimation. The omitted variable has an implicit brand-size constant of zero. Let

$$x^i_{0Kk}(n) = \begin{cases} 1 & \text{if brand-size } k = K, \\ 0 & \text{otherwise.} \end{cases}$$

The correspoding coefficients, b_{0K}, capture any constant uniqueness of an alternative that is not explained by the other variables.

(2) Common across alternatives.

The first common variable is depromoted price, i.e., the regular price in the absence of store promotion.

$x^i_{1k}(n)$ = regular (depromoted) price of brand-size k at the time of customer i 's nth coffee purchase, divided by average category price during week of purchase, each expressed in dollars/ounce.

The second variable represents store promotion,

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

The third variable, depth of promotional price cut, helps explain the differences in response for different promotions.

$x^i_{3k}(n)$ = promotional price cut on brand-size k at time of customer i's nth coffee purchase, divided by average category price during week of purchase.

The variable is zero when there is no promotion on k so that the actual price of brand-size k is always $x_{1k}+x_{3k}$. In our previous paper we did not normalize by category price. We do so now because we subsequently wish to use category price to help explain the "buy now or later" decision. Permitting category price to enter implicitly here through the brand prices would diminish the effect later.

Another difference from the model of our earlier paper is that we are omitting variables relating to when a customer's previous and second previous purchases were on promotion. This is done for simplicity and has no appreciable effect on measures of fit.

A final subset of variables depends on characteristics of the customer. We use the word loyalty to describe the tendency of customers to repurchase the same brand or size.

$x^i_{4k}(n)$ = brand loyalty for brand of brand-size k for the nth coffee purchase of customer i.

$x^i_{5k}(n)$ = size loyalty for size of brand-size k for the nth coffee purchase of customer i.

We define

$$x^i_{4k}(n) = \alpha_b \, x^i_{4k}(n-1)+(1-\alpha_b) \begin{cases} 1 & \text{if customer i bought brand of alternative k at purchase n-1,} \\ 0 & \text{otherwise.} \end{cases}$$

$$x^i_{5k}(n) = \alpha_s \, x^i_{5k}(n-1)+(1-\alpha_s) \begin{cases} 1 & \text{if customer i bought size of alternative k at purchase n-1,} \\ 0 & \text{otherwise.} \end{cases}$$

To start up brand loyalty we set $x^i_{4k}(0)$ to be α_b if the brand of alternative k was the first purchase of the data history of customer i, otherwise $(1-\alpha_b)$/(number of brands - 1), thus insuring that the sum of loyalties across brands always equals 1 for a customer. Startup for size loyalty is analagous.

Variables x^i_{1k} through x^i_{5k} have coeffiecients b_1 through b_5 respectively, which will be estimated in the calibration process. Notice that the b's do not depend on k; that is, the same values are applied to all brands and sizes. This is part of the parsimony of the model formulation.

5.2 Calibration

The calibration of the product choice level of the model uses purchase data from 100 randomly selected panel households. Of the 78 weeks of data available, the first 25 from September 14, 1978 to March 7, 1979 are used for various initializing purposes such as starting up the loyalty variables. This period contains 718 purchases. The next 32 weeks, from March 8, 1979 to October 17, 1979 are used for calibration. During this period the 100 panelists made 1021 purchses, which form the purchase occasions of the calibration.

The results of the calibration are shown in Table 2. As may be seen the store control variables and customer loyalty variables have highly significant coefficients.

PRODUCT CHOICE MODEL

Variable	Coefficient	t-statistic
Brand loyalty	3.76	21.9
Size loyalty	2.97	15.9
Store promotion	2.04	13.7
Promotional price cut	5.26	7.6
Price (depromoted)	-5.11	-7.1
Maxwell House Small constant	-0.06	-0.4
Maxwell House Large constant (omitted)	0	0
Butternut Small constant	0.08	0.5
Butternut Large constant	-0.12	-0.7
Folgers Small constant	0.40	2.4
Folgers Large constant	0.10	0.6
Folgers Flaked Small constant	0.04	0.1
Mellow Roast Small constant	-1.74	-4.7

n = 1021 log likelihood = -978 U^2 = .48

Table 2. Calibration of the product choice model yields highly significant coefficients for store control variables and customer loyalties and, overall, a good fit: U^2 = .48 relative to a null model of choice probability equals market share.

Because the unit of observation is a purchase, we are combining cross-sectional and time series data. One consequence of this is that a good part of the high t-statistics for coefficients of the loyalty variables arises because they explain much of cross-sectional variability in the population. The reason is that, as has been observed (e.g. Bass, Givon, Kalwani, Reibstein, and Wright 1984), purchase probabilities are fairly stable within household but differ from one household to another. This combination of stability and heterogeniety is intentionaly captured by the loyalty variables, which become excellent predictors of product choice.

To clarify the situation Samuels (1983) and Lattin (1984) break loyalty into a preference variable that can be pre-specified by analyzing a pre-calibration period and a learning or adapting variable defined over the calibration period. Such a decomposition is valuable for interpreting the logit coefficients but does not have a large effect on prediction or on the coefficients of the other explanatory variables. Our data set appears to be too short to build a reliable preference measure in the pre-calibration period. When we have tried to do it, the adapting process in the calibration period appears to represent, in part, the model learning the customer preferences rather than the customers learning new behavior. For these reasons and because the thrust of our effort is toward the category purchase not the brand, we have retained the simpler brand choice formulation of the original paper.

5.3 Testing

The holdout sample of 100 households provides a way to the test the model and its calibration. Figures 4-7 of Guadagni and Little (1983) show various plots of predicted and actual shares of each brand-size by four-week period, applying the calibrated model to the 100 customers of the holdout sample. Tracking is good.

U^2 can be calculated for the holdout sample by using (10) on the reference and final models of the calibration sample. As seen in Table 3, the holdout sample actually has a higher U^2 (.48 goes up to .53).

A further measure of predicted quality is $\bar{P}$, the average predicted probability for the brand-sizes actually purchased. In Table 3 we compare its value for calibration and holdout samples over the calibration period and find practically no deterioration in prediction as we go to fresh data (.53 drops to .52).

	U^2	$\bar{P}$
Calibration sample	.48	.53
Holdout sample	.53	.52

Table 3. The calibrated model performs well on the holdout sample with respect to the aggregate measures U^2 and $\bar{P}$. $\bar{P}$ is the average predicted probability for actual choices and U^2 the fraction of uncertainty explained. The measures are taken over the 32 weeks of the calibration period.

The purchase probabilities evaluated by $\bar{P}$ may be described as one purchase ahead predictions. This is because the household purchase history up to a given point is used to update the loyalties and calculate the next purchase probability. A more stringent test is to simulate the household purchases many weeks into the future by Monte Carlo methods. This was done for the model in our earlier paper. Each of the holdout households was initialized with brand and size loyalties from the 25 week pre-calibration period and followed through the 32 weeks of the calibration period with actual purcheses used in the updates, but thereafter all its purchases were determined by Monte Carlo and then used to update loyalties for subsequent purchase occasions over a 20 week projection period. Predicted and actual comparisons at the individual level can no longer be expected to be meaningful because individual households may reasonably wander quite far from their original starting points. Aggregage comparisons of the overall market representation become the primary interest. As may be seen in the original paper, these are not perfect but certainly quite good.

6. CATEGORY CHOICE: WHEN TO BUY

6.1 Specification of Category Purchase Model

Alternatives. The basic choices at the top of the tree are binary: buy-now or buy-later. These will be denoted: a=1 and a=2, respectively.

Observations. An observation is a category purchase opportunity. For the most part this means a shopping trip, but we need to handle the possibility of multiple purchases on a single trip. We do this by viewing the purchase of each package as a separate decision. For example, if a customer purchases one can of coffee on a shopping trip we would say the trip represents two purchase opportunities: the first when the customer walks into the store, the second immediately after the first purchase. Therefore the total number of purchase opportunities is the number of shopping trips plus the number of packages purchased. The data variable for category purchase, $w^i{}_a(m)$, has values

$$w^i{}_1(m) = \begin{cases} 1 & \text{if customer } i \text{ makes a category purchase on the mth purchase opportunity,} \\ 0 & \text{otherwise.} \end{cases}$$

$$w^i{}_2(m) = \begin{cases} 1 & \text{if customer } i \text{ makes no category purchase on the mth purchase opportunity,} \\ 0 & \text{otherwise.} \end{cases}$$

Since some alternative must be chosen on each purchase opportunity,

$$\sum_a w^i{}_a(m) = 1.$$

Category purchase attribute variables. As in the case of the product choice model, we use a linear function of attribute variables to express the utilities of purchasing in the category now and later. The general form of customer i's utility for alternative a on purchase opportunity m is

$$\sum_j d_j \, z^i{}_{ja}(m) \;.$$

An issue arises because the two alternatives seem quite different: buy-now has hard data associated with it, buy-later is vague and uncertain. Any variable defined for one alternative must have a value for the other (unless the variable is unique to an alternative, in which case its value for the other is defined as zero). The natural tendency is to make most variables unique to one or the other alternative. However, the power of the model is likely to reside in variables that provide relevant comparisons between buy-now and buy-later. Table 4 displays the ones we have devised. Of special interest are household inventory, category attractiveness, and category price.

	Value of variable for	
Explanatory Variable	a=1 Buy now	a=2 Buy later
Buy-now dummy (z_0)	1	0
Purchase opportunity within trip (z_1)		
m = first	1	0
m = later	0	0
Household inventory of coffee (z_2)	estimated inventory	0
Category attractiveness (z_3)	expected max utility for household from choice model	8 week moving average of expected max utility for household
Category price (z_4)	current average category price for household	8 week moving average of average category price for household
Freeze announcement (z_5)	1	0

Table 4. Strategy for handling buy-now and buy-later values for category choice variables.

Buy-now dummy. The utility of the buy-now alternative includes an additive constant specific to the alternative. The form is a dummy variable:

$$z^i_{01}(m) = \begin{cases} 1 & \text{for buy-now,} \\ 0 & \text{otherwise,} \end{cases}$$

$$z^i_{02}(m) = 0$$

with coefficient, d_0. The alternative specific constant for the buy-later alternative is implicitly zero. These constants serve to express any uniqueness of the two alternatives across all observations that is not captured by other explanatory variables.

First purchase opportunity dummy. Another variable specific to the buy-now alternative is a dummy to differentiate the first purchase opportunity on a shopping trip from later ones. Its coefficient, d_1, will capture the increase or decrease of the likelihood of a purchase, given that it is the first purchase of the trip, insofar as this is not explained by other variables.

$$z^{i}_{11}(m) = \begin{cases} 1 & \text{if customer i's mth purchase opportunity is the first purchase opportunity of a shopping trip,} \\ 0 & \text{otherwise.} \end{cases}$$

$$z^{i}_{12}(m) = 0.$$

<u>Household inventory</u>. The amount of coffee the household has on hand obviously drives its decision of whether to buy now or later and so deserves careful treatment. We wish to estimate household inventory of coffee at the time of each purchase opportunity. To adjust for diffences in consumption rates across households, inventory is measured in weeks of supply. The form of the inventory term is

$$z^{i}_{21}(m) = z^{i}_{21}(m-1)-(t^{i}(m)-t^{i}(m-1))s(t^{i}(m))+q^{i}(m-1)/c^{i} \quad \text{(buy now)}$$

$$z^{i}_{22}(m) = 0 \qquad \text{(buy later)},$$

where: $q^{i}(m)$ = quantity of coffee purchased by customer i on i's mth purchase opportunity (ounces),

c^{i} = i's average consumption rate (ounces/week).

$t^{i}(m)$ = point in time of customer i's mth purchase opportunity (weeks).

$s(t)$ = seasonal index of coffee consumption.

In words, the household's inventory at purchase opportunity m is its value at the previous opportunity augmented by any purchases and decremented by estimated consumption between opportunities. The calculation of consumption rate is based on the household's purchase history. Consumption is assumed to have a seasonal component, which is taken to have an index that is unity plus a sine curve with amplitude 0.1 peaking January 31 and bottoming July 31.

Note that the construction of the inventory term allows immediate updating. Each coffee purchase on a shopping trip therefore increases inventory prior to the decision whether to buy another package on that trip.

The starting value for a customer's inventory is found by setting it temporarily to zero and calculating a trial inventory throughout the initialization period using the additions and depletions appropriate for the period. We then find the most negative value and take its absolute magnitude as the starting inventory for that customer. If the household's inventory never goes negative in the period, we take the beginning inventory to be zero. This process assumes the low point of inventory to be zero and appears to ignore the possiblity that a household may maintain a buffer stock or threshold to trigger purchase. However, our formulation will be consistent if we correspondingly use zero for the buy-later alternative, since the logit will cancel out any constant buffer added to both alternatives, even if it differs household by household.

Accordingly, the value zero is applied to inventory in the buy-later alternative to represent the notion that the household expects to purchase when its stock reaches a trigger level, which may be unique to the household.

Finally, for robustness against abnormally low consumption periods (e.g., vacations) and extra high consumption periods (e.g., house guests), we clip inventory if it starts to rise above 10 weeks supply or drop below -1 week supply (the latter being considered possible because of buffer stock).

For the next two variables, the value of the buy-later alternative represents the customer's expectation of what will be available at a later date. Since the primary information available to the customer is what has happened in the past, we assume that households will use previous experience to generate expectations for the future. The model does this by taking the value of the buy-later attributes to be the average of the buy-now attributes over the previous 8 weeks.

Category attractiveness. A particularly important variable is the attractiveness of the category as a whole at the purchase opportunity. It is the "inclusive value" or the expected maximum utility of a product choice, as determined from the products available to the customer and their individual utilities in the product choice model. As discussed earlier, the variable equals the log of the denominator of the product choice probability.

$$z^i{}_{31}(m) = \ln \{\sum_j \exp(v^i{}_j(m))\}$$

$$z^i{}_{32}(m) = \sum_{q \epsilon Q^i} z^i{}_{31}(q)/n^i(m)$$

where Q^i is the set of purchase opportunities for customer i in the previous eight weeks before purchase m and n^i is their total number.

Marketing activities, such as promotion, that increase utilities for individual brands increase the value of z_{31} and so the probability of buying now rather than later. Note that, because of the multiplicative effect of adding terms to an exponent, a promotion on a brand for which a customer has high loyalty will produce a particularly strong push for buying the category now.

Category price. The next variable is the category price level at the time of the purchase opportunity:

$$z^i{}_{41}(m) = \sum_k x'^i{}_{1k}(m)/N(m)$$

$$z^i{}_{42}(m) = \sum_{q \epsilon Q^i} z^i{}_{41}(q)/n^i(m) ,$$

where $x'^i{}_{1k}(m)$ is the price of brandsize k in dollars/ounce at the time of purchase opportunity m and N(m) is the number of brand-sizes available. This variable captures the effect of short-term category price fluctuations on coffee purchasing behavior. It presupposes that after some period of time, taken as eight weeks, people will adjust to a new price level.

Freeze dummy. A final variable is designed to capture the effects of a coffee freeze in Brazil. During the time covered in the analysis a major freeze took place in the coffee growing regions of that country. The event was widely reported in the media, and, as is evident in Figure 1, affected the coffee buying public. It seems clear that the publicity raised the specter of higher prices and prompted many households to stock up in the truest sense of buying now rather than later. We operationalize the effect with a dummy variable for the buy-now category purchase alternative:

$$z^i_{51}(m) = \begin{cases} 1 & \text{if i's mth purchase opportunity took place between June 6 and July 4, 1979.} \\ 0 & \text{otherwise.} \end{cases}$$

$$z^i_{52}(m) = 0.$$

Throughout this section we have motivated the category level variables with the choice alternatives, buy-now or buy-later. This has the drawback that the buy-later alternative is not filled out with an actual brand-size choice at the next level. Another, possibly more orthodox formulation is to think about the category decision in terms of "buy" vs. "don't buy" at each purchase opportunity. All variables are unique to "buy". However, we define them to be the difference between the buy-now and buy-later variables presented above. The "don't buy" variables are zero and "don't buy" has no second level in the decision tree. It is relatively easy to verify that, because of the form of the binary logit, the two formulations are mathematically the same. In the new view, the relevant category price for "buy", as an example, would be the difference between current price and a reference price consisting of the eight-week moving average of past values. Similar remarks hold for household inventory, category attractiveness, and, in an equally valid but less interesting way, the other variables.

6.2 Calibration

The category choice model is intialized and calibrated on the same households and time periods as the the product choice model. During the 32 week calibration period the 100 households have 3808 purchase opportunities (observations) consisting of 2787 shopping trips and 1021 category purchases, each of which creates another purchase opportunity. In the 25 week initialization period there are 1758 shopping trips, 718 category purchases, and so 2476 purchase opportunites.

Proceeding in nested-logit fashion, we use the calibrated choice model to calculate z^i_{3a}, the category attractiveness variable. The others come from original data. The explanatory variables combined with the choice observations, w^i_a, comprise the input data to a standard logit parameter estimation program. The results, arrayed in a series of model specifications S1 through S5, are displayed in Table 5.

CATEGORY CHOICE MODEL

Specification	S1	S2	S3	S4	S5
U^2	0	.032	.052	.053	.056
log likelihood	-2,052	-1,986	-1,945	-1,944	-1,937
Buy-now dummy	-2.52 (-21.3)	-1.97 (-15.5)	-2.09 (-16.2)	-2.10 (-16.2)	-2.15 (-16.5)
Buy on first purchase opportunity of trip	1.91 (15.3)	1.66 (13.0)	1.77 (13.7)	1.78 (13.8)	1.78 (13.8)
Household inventory		-.164 (-10.9)	-.164 (-10.8)	-.162 (-10.6)	-.170 (-11.0)
Category attractiveness			.099 (8.8)	.096 (8.4)	.091 (7.9)
Category price				-4.86 (-1.1)	-4.34 (-0.9)
Freeze dummy					.440 (3.9)

Table 5. Calibration of the category choice model for increasing numbers of explanatory variables displays highly significant coefficients for the key variables household inventory and category attractiveness.

The final model is built up by adding variables to a basic specification, S1, which serves as a reference model. This process brings out the contribution of the individual variables and also demonstrates that the coefficients are quite stable across specifications, once a few important attributes are present. Specification S1 contains only the "buy-now" and "buy on first purchase opportunity" dummies. Basically, it plays back that 35% of the trips lead to a coffee purchase and forms our reference model for measuring improvements.

Household coffee inventory has a powerful influence on purchase. In model S2, if a household has on hand a four-week supply of coffee, its probability of making a purchase on a shopping trip is .28. If, on the other hand, the household has a one-week supply, the probability jumps to .39, an increase of 40%.

Category attractiveness is almost as important, having a coefficient with high statistical significance. The variable is primarily influenced by promotion. Using specification S3 and illustrative data from our sample, we find that a difference of 4 units between the value of category attractiveness and its average of the previous eight weeks changes a .38 probability of purchase into .48, for a 26% increase.

Category price, on the other hand, does not show much effect. Recall that category price is the average price/ounce across brand-sizes during the week of purchase and is being compared to the previous eight week average of the same variable. A 10% increase in category price would produce only about a 5% decrease in probability of purchase for a household with one week of inventory on hand.

The insensitivity of sales to category price occurs despite a big swing in that variable due to the freeze in Brazil. In our data, coffee sales actually shoot up while price stays flat, but this is accounted for by the freeze dummy. Later on, price rises and sales fall, but the model calibration finds that this phenomenon is more easily explained by high household inventories than by category price. Thus, at least for our model and data, the relatively small changes in category price over most of the time period are not especially related to changes in category purchase probability.

On the other hand, the freeze dummy that models the media-publicized threat of a price increase accounts for a substantial jump in category purchases. Using S5, the probability that a household with a one week coffee supply will purchase is .37. With the freeze news, it jumps to .48. Although this may not seem to be a dramatic change for a single household, it will produce a big surge for the stores and manufacturers. The model does not let the bump last long. Rising household inventories quickly deter further loading. A real category price increase signals the end of anticipation and creates a gap compared to previous experience for a further, though less important, negative effect on sales.

6.3 Testing and Evaluation

We shall evaluate final model, S5, three ways by examining: (1) U^2, the degree of uncertainty explained, (2) P, the average probability of the alternative actually chosen, and (3) plots comparing predicted and actual purchases over time.

U-Squared. U^2 measures the amount of uncertainty explained relative to a reference model. The reference model is S1, which plays back the percentage of trips on which a coffee purchase was made. How much do we increase U^2 beyond that as a base? A little, but not much: $U^2 = .056$ for the final model of the calibration sample. Table 5 shows that most of the contribution comes from category attractiveness and inventory. Likelihood ratio tests establish that the addition of each of these parameters to the model is significant at high levels (well beyond .001). The addition of the freeze dummy is also significant but category price is not. We leave the latter in the model for completeness.

The value of U^2 seems small compared to the .48 attained in the brand choice model. One reason is that our category reference model is already a good predictor and so reduces the room for improvement. The category reference model says that the probability is .65 for choosing the buy-later alternative when the customer walks into the store. By way of contrast, the largest choice probability in the brand reference model is that for Folgers Small at .38. Another reason for the higher U^2 in the brand model is the extreme power of the loyalty and promotion variables for predicting choice. They are better than inventory and category attractiveness, even though these are very good.

$\bar{P}$, average probability of chosen alternative. Table 6 shows $\bar{P}$, the average probability of the alternative actually chosen, for calibration and holdout samples, as calculated by the model developed on the calibration sample. The deterioration is very small, from .66 to .65, indicating that the individual choice probabilities are predicted just as well in the holdout sample as in the calibration data.

	$\bar{P}$
Calibration sample	.66
Holdout sample	.65

Table 6. The category choice model developed on the calibration sample performs well on the holdout sample with respect to P, the average probability for the category alternative actually selected.

Plots of predicted and actual. Plots of predicted and actual purchases by week over the calibration period for both calibration and holdout samples appear in Figure 3. To provide a measure of sampling variation attributable to the relatively small number of purchases in each week, we have plotted approximate 90% confidence intervals. (These assume each purchase opportunity to be an independent binomial draw with probability given by the model. This calculation probably underestimates the sizes of the intervals somewhat. See Guadagni and Little (1983).)

We find that 3 of the 32 weeks lie outside the confidence interval in the calibration sample. This is about what would be expected if the predicted and actual sales were the same to within sampling tolerances. In the holdout sample 7 of 32 weeks lie outside. This is a larger number than is likely to be due to chance and so indicates that predicted differs from actual. Thus there is some deterioration of tracking performance.

(Note that if we had enough data so that the confidence intervals could be very tight, almost all the points would fall outside in both cases, since predicted is bound to be at least a little different from actual. The purpose of the confidence intervals is to prevent overinterpretation of deviations when they are simply due to small numbers of purchases.)

In any case, although the holdout sample does not fit quite as well

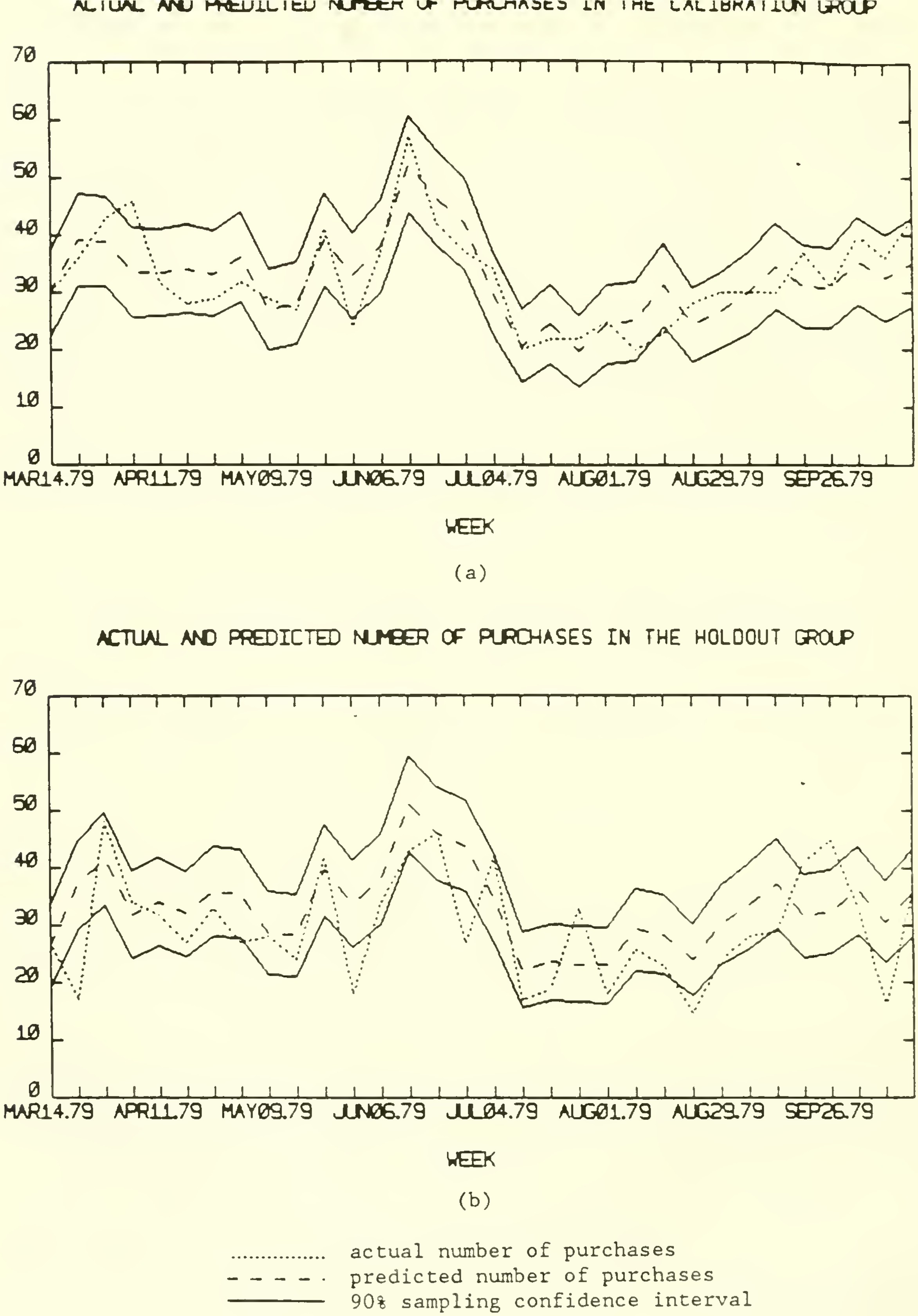


Figure 3. Plots of predicted and actual coffee purchases for (a) the calibration sample compared to (b) the holdout sample show some deterioration in performance but overall are quite good.

as the calibration sample, the main features of the aggregate coffee market remain evident, namely, the build-up of sales in June, a lull in August, and the recovery in September.

4. COMBINED MODEL

The final step is to put the two submodels together into a single predictor of brand-size purchases, household by household, shopping trip by shopping trip. Individual probabilities then aggregate to market level sales and permit calculations of how the market responds to control variables.

Tracking quality. We skip the calibration sample and go straight to the holdout. Figure 4 shows predicted and actual purchases for the eight brand-sizes. The time period includes not only the weeks of the calibration sample but also a further forecast period starting October 18, 1979, separated in each plot by a vertical line.

Comparison of these plots with the share curves in our earlier paper shows that the new figures are noticeably different, the reason being that category level phenomena (e.g. the Brazil freeze) are now affecting brand-size sales.

The quality of the tracking is quite good, even into the forecast period. The major features of actual sales for most brand-sizes are captured well by the model, especially considering the many twists and turns that the sales of some products go through as their marketing activities change. However, the fit generally is not quite as good as in the share model alone. This may not be surprising, since we are asking the model to do much more than before. As might be expected, places where the tracking was weak in the share model, for example, a period of over-prediction in Butternut Small and the miss on the level of Folgers Flaked continue to be weak and, in fact, are somewhat exacerbated.

The overprediction of Butternut Small is the result of two promotional events that did not boost sales as much as the model predicted. Note that the model handles all store promotions for all brands with only two parameters. Since it seems unlikely that all store promotions are really equally effective, some deviations of this kind (both up and down) are to be anticipated. Considering the parsimony of the model, the overall tracking seem satisfying.

Application to market response analysis. An important practical reason for modeling category purchase is to enable a better calculation of sales response. This is possible because the combined model considers the effect of marketing actions on total sales and not just share. To illustrate the point, Table 7 displays the short run sales response to in-store promotion for each brand-size based on the combined model and, for comparison, the corresponding share response from our earlier paper. As may be seen, sales shows a higher response than share, the reason being that promotion expands category sales in the short run. (The methodology for performing the calculations appears in our previous paper.)

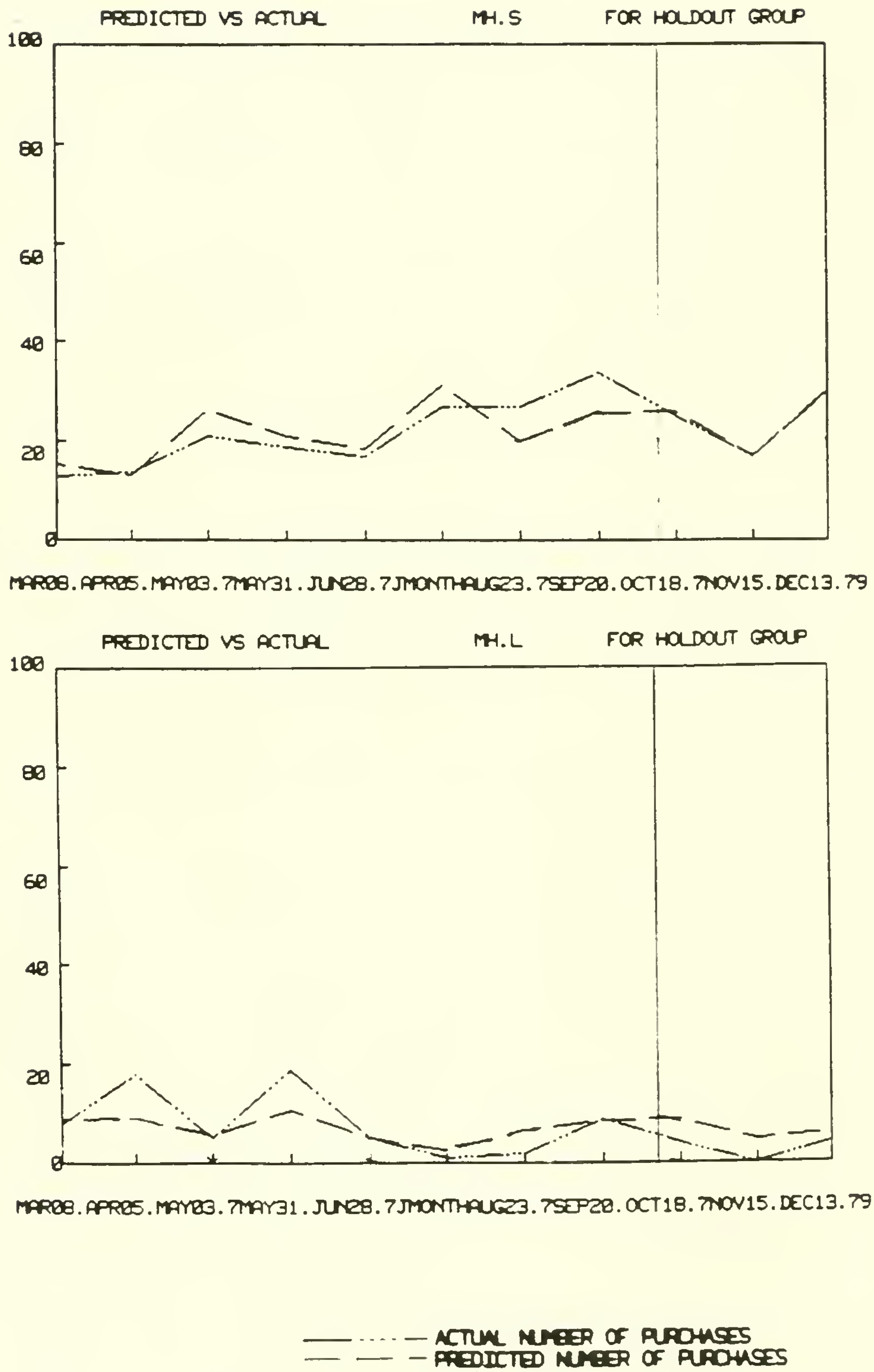
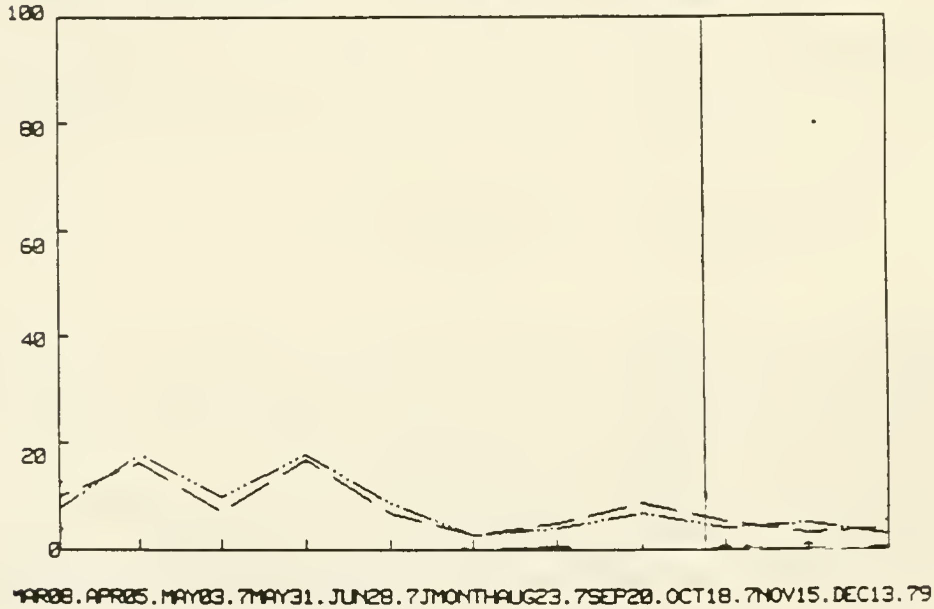


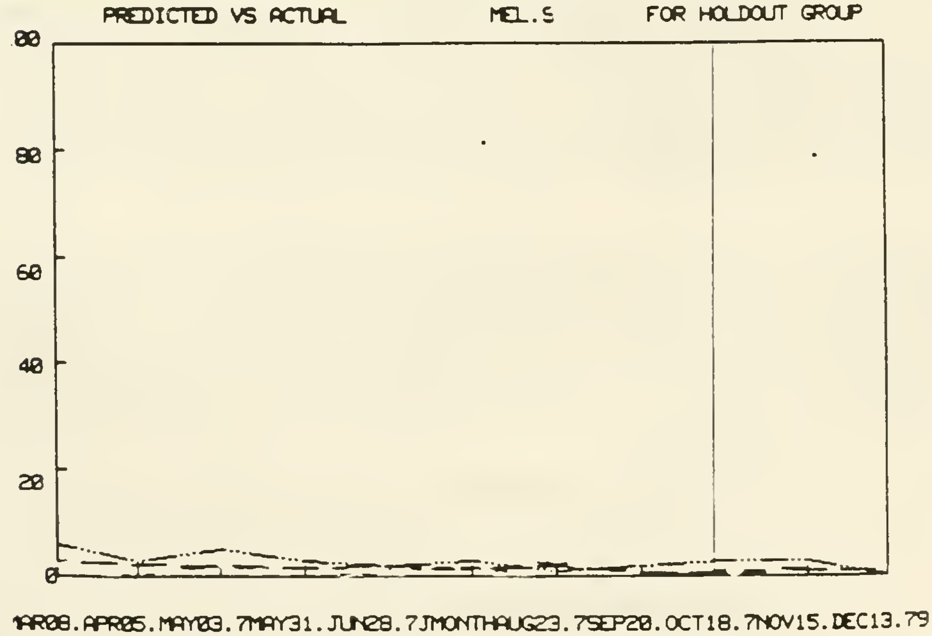
Figure 4. Predicted and actual sales (in packages) using the combined model on the holdout sample show generally good agreement, although there are some deviations in level. Turning points are tracked quite well.

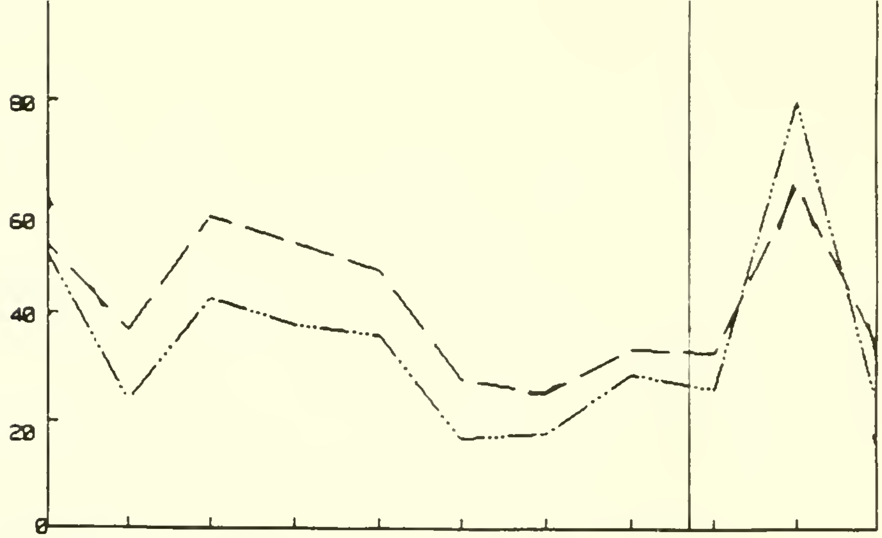
PREDICTED VS ACTUAL BUT.S FOR HOLDOUT GROUP

PREDICTED VS ACTUAL BUT.L FOR HOLDOUT GROUP

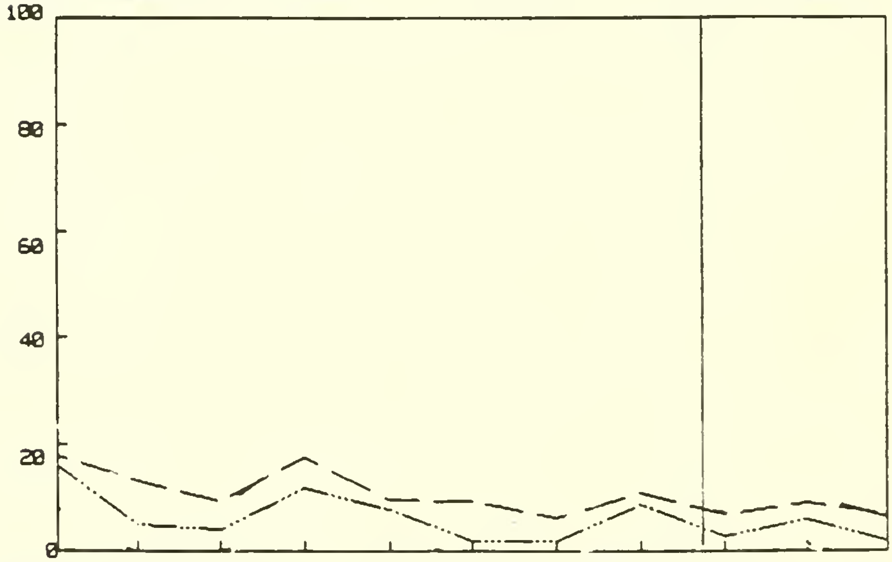


PREDICTED VS ACTUAL MEL.S FOR HOLDOUT GROUP

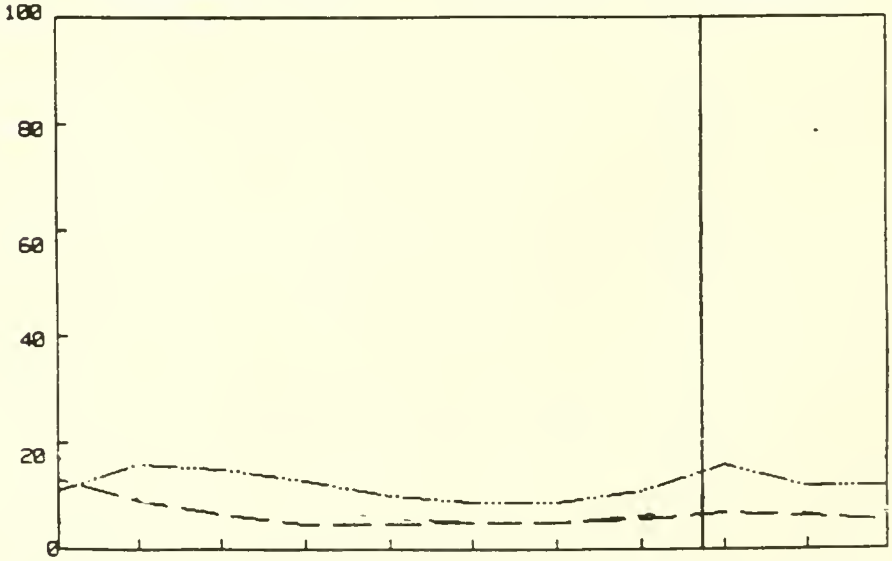




PREDICTED VS ACTUAL FOL.L FOR HOLDOUT GROUP



PREDICTED VS ACTUAL FOL.F.S FOR HOLDOUT GROUP



	Response to a promotion with a median price cut	
	sales (%)	share (%)
Folgers Small	315	173
Butternut Small	508	334
Folgers Large	450	273
Maxwell House Small	614	362
Maxwell House Large	534	363
Butternut Large	813	502
Folgers Flaked Small	721	409
Mellow Roast Small	924	568

Table 7. Short run response to in-store promotion using the combined model shows that sales response exceeds the share response calculated in Guadagni and Little (1983) by 50% to 80%.

5. CONCLUSIONS

What can we conclude? Our earlier paper demonstrated that household level logit models are powerful for identifying and measuring the effect of marketing variables on product choice and that the calculations required are quite feasible. The present paper shows that the nested logit and its sequential estimation does an effective job at bringing in category level effects based on shopping trips. The computations are again feasible and can be done with readily available computer packages. Sequential estimation is not as efficient as a full maximum likelihood calibration, however, and the latter deserves further investigation.

The multinomial logit, in general, and the nested logit, in particular, make a variety of strong assumptions, particularly about error terms. To provide as much protection as possible against false inferences, we have tested our models on holdout samples and in forecast periods. Results are not perfect but certainly stand up quite well to the several types of evaluation.

The category choice model shows that several theoretically-motivated variables are important in determining coffee purchase on a shopping trip, especially household inventory and category attractiveness. The latter comes out of the product choice model and represents the expected maximum utility of the brand-sizes available to the customer on the trip. This is strongly affected by in-store promotion.

Finally, the combined model, incorporating both category and product choice, permits a better, if more elaborate, calculation of sales response to marketing variables than does the product choice model alone. We have illustrated this by calculating the short run response to store promotion. A higher response would be expected and is found.

REFERENCES

Bass, F.M., M.M. Givon, M.U. Kalwani, D. Reibstein, and G.P. Wright (1984), "An Investigation into the Order of the Brand Choice Process," Marketing Science, Vol. 3, No. 4, Fall 1984.

Ben Akiva, M. and S. Lerman (1985), Discrete Choice Analysis, M.I.T. Press, Cambridge, MA 02139.

Guadagni, P.M. and J.D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," Marketing Science, Vol. 2, No. 3, Summer 1983.

Guadagni, P.M. (1983), "A Nested Logit Model of Product Choice and Purchase Incidence," Proceedings of the 1983 Marketing Science Conference, TIMS, Providence, RI.

Lattin, J. (1984), Ph.D. thesis, M.I.T. Sloan School, Cambridge, MA 02139.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden, Structural Analysis of Discrete Data with Econometric Applications, M.I.T. Press, Cambridge, MA 02139.

Samuels, D. (1983), S.M. thesis, M.I.T. Sloan School, Cambridge, MA 02139.