European Journal of Operational Research 76 (1994) 309–320
North-Holland


# Implications and relative fit of several first-order Markov models of consumer variety seeking

Fred M. Feinberg
*The Fuqua School of Business, Duke University, Durham, NC 27706, USA*

Barbara E. Kahn
*University of Toronto, Toronto, Ont., Canada*

Leigh McAlister
*University of Texas, Austin, TX, USA*

Revised March 1991

**Abstract:** We derive $n$-brand solutions for several first-order Markov models of consumer variety seeking, among them those of Givon (1984), Lattin and McAlister (1985) and several variants. Such solutions allow a comparative static analysis of long-run market shares relative to changes in the models' parameters: variety-seeking intensity, brand preference and degree of feature sharing. Along with a simple Multinomial and a fully general first-order Markov model, these models are calibrated for consumers in a behavioral experiment. Such fits allow the nested models to be compared through a likelihood ratio test, and the non-nested ones to be compared through Hauser's (1978) $U^2$ measure. These tests indicate that the Lattin–McAlister model performs arguably better than all but the general Markov model.

**Keywords:** Variety-seeking; Brand choice models; Markov models

## Introduction

Since the pioneering work of Bass, Pessemier and Lehmann (1972), the pervasive effects of consumer variety seeking have drawn increasing attention. In their thorough review of the extant literature, McAlister and Pessemier (1982) suggest that consumers can behave in a way that suggests a desire for change in and of itself; moreover, this behavior is distinct from the brand-switching that arises whenever a consumer chooses among the set of brands for which he has preference. It is possible to interpret this distinction as one of the order of the brand choice process, contrasting the ostensibly zero-order process of normal-brand switching with the possibly higher-order process of switching with an explicit desire for change. Bass et al. (1984) discuss this issue in depth, pointing out the difficulties in determining the order of the choice process from a particular string of purchases, and how this can often be confounded with non-stationarity of preference. They explore a variety

*Correspondence to:* Prof. F.M. Feinberg, University of Toronto, Toronto. Ont., Canada.


*SSDI* 0377-2217(93)E0304-G

of statistical methods to clarify these issues, and offer evidence that both zero-order and first-order purchase behavior occur in several product categories.

Several papers have appeared which distinguish zero- and first-order behavior by modeling brand-switching behavior as a Markov process. Often, due to parsimony, it is necessary to restrict analysis to a two-brand framework, looking at the brand-switching process as taking place between a target brand, perhaps the favorite brand, and a category composed of all others. Some of the earliest models, those of Morrison (1966) and Jones (1973), specify different parameterizations of the Markov brand-switching matrix without directly tying these parameters to managerial control variables. Jeuland (1979) extended Morrison's work to include a specific parameter measuring inertia or loyal behavior. Givon (1984) proposed a more general model, in which the intensity of variety-seeking, in addition to Jeuland's variety-avoiding, is explicitly factored into the parameterization. A further step in this direction was the model of Lattin and McAlister (1985), which included not only the consumer-specific level of variety-seeking, but brand preference and interbrand similarity variables as well, imbedded in an $n$-brand choice set context. For a full discussion of feature-based models of brand switching, see Hutchinson (1986).

Because of its feature-based orientation, it is possible to use the Lattin–McAlister model to predict likely changes in long-run purchase probabilities relative to changes in parameters under managerial influence: variety-seeking intensity, brand preference and perceptions of inter-brand similarity; for the Givon model, this can be done for the variety-seeking parameter only; for a Multinomial model, no such analysis is possible.

We focus our attention on six Markov models of consumer variety seeking: the simple Multinomial model; the Givon model and a variant accounting for consistent conditional purchase probabilities; the Lattin–McAlister model and a parsimonious variant; and the completely unrestricted first-order Markov model. We derive the explicit solution to these models in the somewhat restrictive two-brand case, and matrix-based solutions and their derivatives in the more complex $n$-brand case. These $n$-brand solutions are used in a 3-brand simulation for a thorough static analysis. Finally, the models are fitted to the shampoo choice histories of consumers in a computer experiment, and these fits compared, in a maximum-likelihood setting, to those obtainable from other first-order model formulations.

## Modeling framework

The essential feature of Markov models of brand-switching is a set of *transition probabilities*, $a_{ji}$, each stating the probability of choosing a particular brand $B_i$ directly following the selection of $B_j$. Such models directly imply that selection on one choice occasion affects preferences on the next choice occasion. Having just consumed a particular bundle of features (i.e., a brand), a variety seeker is less attracted to those features when making the next choice; therefore, brands composed of different features become relatively more desirable. We refer to a consumer whose known choice set consists of $n$ brands, denoted $\{B_i\}$. Several parametric conventions are staples of the models we will consider here. For consistency in discussion, we use the following notation:

$U_j$ = The value to the consumer of $B_j$'s unique features.

$S_{ij}$ = The value to the consumer of all features shared by brands $B_i$ and $B_j$. By definition, $S_{ij} = S_{ji}$, and $S_{jj} = \pi_j$. For notational convenience, define $S_{\cdot j} = \sum_k S_{kj}$.

$\pi_j$ = The consumer's unconditional preference for $B_j$, representing the total value to this consumer of all features – unique and shared – provided by $B_j$, scaled so that $\sum_j \pi_j = 1$. By definition

$$\pi_j = U_j + \sum_{j \neq k} S_{kj} \quad \text{for any } j.$$

$V$ = A discount factor indicating the variety seeking intensity of the consumer; $0 \le V \le 1$.

$x_j$ = The steady-state market share for $B_j$ resulting from the Markov model.

It is possible to compare $x_j$, the steady-state market share for $B_j$ implied by the first order Markov choice process, and $\pi_j$, the unconditional preference share for $B_j$, concluding that variety-seeking is beneficial only when $x_j - \pi_j > 0$, as this indicates $B_j$ is chosen more than it would be based on its zero order preference, $\pi_j$. [1] Within the framework of this parameterization, the size of the observed market share, $x_j$, might be influenced as follows:

- Changing $V$ by influencing the intensity of variety seeking behavior $(\partial x_j/\partial V)$;
- Changing $\pi_j$ by altering the value of a brand's unique features $(\partial x_j/\partial U_j)$;
- Changing $\pi_j$ by altering the value of a brand's unique features $(\partial x_j/\partial S_{ij})$.

Since the explicit (non-matrix) calculation of $x_j$ and appropriate derivatives is only reasonable when there are just two brands, we will, after a brief discussion of the models in question, first turn our attention there. Bass et. al. (1984) discuss the importance of, and tests germane to, this case, essentially simplifying a Multinomial process through a related Binomial, comparing runs of a 'favorite brand' against a background of purchases of all other brands.

## Model formulations

We examine the following models, listed in descending order of number of estimated parameters:

*Unrestricted Markov:* This model specifies the entire set of $n^2$ transition probabilities, $a_{ji}$, which are fitted directly. Since it is fully general and unrestricted (outside of its rows summing to one), it will fit at least as well as any other Markov model. There are $n(n-1)$ estimated parameters.

*Lattin–McAlister:* A feature-based model, preference for brand $B_i$ is decomposed into its unique features $(U_i)$ and features it shares with other brands $(S_{ij})$. The more features they share, the less likely brand $B_i$ is to be chosen after brand $B_j$. Transition probabilities are given by

$$a_{ji} = (\pi_i - VS_{ij})/(1 - VS_{\cdot j}).$$

There are $n-1$ independent unique feature parameters $(U_i)$, $\frac{1}{2}n(n-1)$ shared feature parameters, and a single variety-seeking parameter; this implies $\frac{1}{2}n(n+1)$ total estimated parameters.

*Restricted Lattin–McAlister:* To ensure a fairer comparison of the Lattin–McAlister model and more parsimonious models, it is possible to reduce the number of estimated parameters. This can be done by assuming that all brands 'overlap' by an expected amount, probabilistically, setting $S_{ij} = \pi_i\pi_j$. This reduces the number of estimated parameters to $n$.

*Givon (1984):* To specify transition probabilities, the model linearly combines a brand's preference $(\pi_i)$ and a fixed propensity to switch to another brand $(V)$, irrespective of the preference of the brand switched to, so long as it is different from the brand switched from. Transition probabilities are thus given by

$$a_{ji} = (1 - V)\pi_i + V/(n-1)$$

for $i$ different from $j$, and

$$a_{ii} = (1 - V)\pi_i.$$

Fitted parameters are $\{V, \pi_1, \ldots, \pi_n\}$, with $n$ estimated parameters.

*Modified Givon:* It is possible to modify the Givon model to account for the greater likelihood of switching to a more preferred brand than to a less preferred one. Rather than distribute $V$ equally amongst all other brands, it is possible to do so in proportion to their preference. Transition probabilities are then given by

$$a_{ji} = (1 - V)\pi_i + (V)\pi_i/(1 - \pi_j)$$

for $i$ different from $j$, and

$$a_{ii} = (1 - V)\pi_i.$$

Fitted parameters are still $\{V, \pi_1, \ldots, \pi_n\}$, with $n$ estimated parameters.

*Multinomial:* The standard Multinomial model suggests that there is no dependence among choices over time. It is thus equivalent to the unrestricted Markov model with each row of transition probabilities equal, or any of the other models with $V$ set to zero. Since the conditional probability of choosing $B_i$ following $B_j$ is independent of $B_j$, $a_{ji} = \pi_i$. Fitted parameters are $\{\pi_1, \ldots, \pi_n\}$, with $n-1$ estimated parameters.

While it is possible to fit a Binomial model as well, as this is anticipated to fit far less well than

[1] Note that, in Jeuland's inertia model (1979), the steady-state market share for the brand is equal to the unconditional preference for the brand. This is not true of variety-seeking models in general, where the probability of switching to a brand deviates from its unconditional preference toward an equal probability distribution across all brands not purchased on the last purchase occasion. For a discussion of this point, see Givon (1984).

even a simple Multinomial model, we have omitted it here. In the next section, we present two-brand analyses of the models under parametric variations.

## Two brand case

Given that a particular consumer's consideration set consists of two brands, $B_1$ and $B_2$, we define the four transition probabilities:

$a_{11} = \text{Prob(repeat consumption } B_1)$,

$a_{22} = \text{Prob(repeat consumption } B_2)$,

$a_{12} = \text{Prob(consume } B_2 \text{ after having consumed } B_1)$,

and

$a_{21} = \text{Prob(consume } B_1 \text{ after having consumed } B_2)$.

Note that

$$a_{11} + a_{12} = 1 \quad \text{and} \quad a_{21} + a_{22} = 1.$$

We have two equations in the unknowns $x_1$ and $x_2$, a linear system of the form $X = XA$, where $X$ is the row vector $[x_1 x_2]$, of steady-state market shares for $B_1$ and $B_2$. This a Markov process with transition matrix $A$ given by

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The expected market share, or steady-state choice probability, for $B_1$ is

$$x_1 = a_{21}/(a_{21} + a_{12}). \tag{1}$$

To understand how parametric changes influence market share, we need to compute $(\partial x_i/\partial \eta)$, where $\eta = V$, $\pi_i$, or $S_{ij}$, the last of these applicable only to the Lattin–McAlister based models.

## Market share, preference share and changing parameters

Were there no variety seeking ($V = 0$), we would expect a brand's market share and preference share to be equal. Since variety seeking can cause a share redistribution, we focus on the differential between steady-state share and underlying preference, $x_1 - \pi_1$. Further, to examine the parametric effects of the different models in the simple two-brand case, we examine the derivatives $(\partial x_1/\partial V)$, $(\partial x_1/\partial U_1)$ and $(\partial x_1/\partial S_{12})$. Throughout, we use the convention that brands are numbered in descending order of preference (e.g., $\pi_1 \geq \pi_2$). Since $x_1$ and $x_2$ are fixed to sum to one, no generality is lost by focusing on $B_1$, the more preferred brand.

Notice that, when there are two brands, the Givon and Modified Givon models are equivalent. Using (1), we have

$$x_1 - \pi_1 = \frac{-(\pi_1 - \pi_2)V}{1 - V} \leq 0, \tag{2a}$$

$$\frac{\partial x_1}{\partial V} = \frac{-(\pi_1 - \pi_2)}{(1 - V)^2} \leq 0, \tag{2b}$$

$$\frac{\partial x_1}{\partial \pi_1} = \frac{1 - V}{1 + V} \geq 0. \tag{2c}$$

Thus, for these models, there is a basic disposition to have variety-seeking penalize larger preference brands overall ($x_1 - \pi_1 \leq 0$) and to have the degree of penalization increase with the level of variety-seeking intensity $((\partial x_1/\partial V) \leq 0)$. As can be expected, an increase in the level of brand preference translates to an increase in share, even after variety-seeking is accounted for $((\partial x_1/\partial \pi_1) \geq 0)$.

For the Lattin–McAlister model, we have the following (note: $f(\cdot) = [(\pi_1 - VS_{12})(1 - VS_{\cdot 1}) + (\pi_2 - VS_{12})(1 - VS_{\cdot 2})]$, as a positive divisor of all quantities here, has no effect on sign; rules for differentiation are discussed in the Appendix):

$$x_1 - \pi_1 = V(\pi_1 - \pi_2)\frac{[S_{12}(1 - VS_{12}) - \pi_1\pi_2]}{f(\cdot)}, \tag{3a}$$

$$\frac{\partial x_1}{\partial V} = (\pi_1 - \pi_2) \times \frac{[S_{12}(1 - VS_{12})^2 - (1 - V^2 S_{12})\pi_1\pi_2]}{f^2(\cdot)}, \tag{3b}$$

$$\frac{\partial x_1}{\partial U_1} = (1 - V)(\pi_2 - VS_{12}) \times \frac{[(1 - VS_{\cdot 2})(1 - VS_{12}) - V\pi_2(\pi_1 - \pi_2)]}{f^2(\cdot)}, \tag{3c}$$

$$\frac{\partial x_1}{\partial S_{12}} = (\pi_1 - \pi_2)V(1-V) \times \frac{\left[(1-VS_{12})^2 - V\pi_1\pi_2\right]}{f^2(\cdot)}. \qquad (3d)$$

Notice that, if $S_{12} \le \pi_1\pi_2$ (the smaller preference brand does not derive 'too much' of its value from shared features, as in the Restricted Lattin–McAlister model), $x_1 - \pi_1$ is negative, indicating that the smaller brand will gain share and the larger preference brand lose share. In situations with intense variety seeking ($V$ near 1), the smaller preference brand can derive almost of all of its value from shared features ($S_{12}$ near $\min\{\pi_1, \pi_2\}$) and still enjoy a market share that exceeds its preference share.

It is not difficult to show [2] that, in a two-brand choice set, (3a–d) imply that $(\partial x_1/\partial V)$ is negative when $S_{12} < \pi_1\pi_2$ and that both $(\partial x_1/\partial U_1)$ and $(\partial x_1/\partial S_{12})$ are always positive (recall $\pi_1 > \pi_2$). This pivotal condition, $S_{12} = \pi_1\pi_2$, is precisely the one used to define the Restricted Lattin–McAlister model. Bearing in mind the caveat that the two-brand case ignores the possibly complex interactions of additional brands, we can use these derivatives to foreshadow several tendencies of these models:

*Increasing variety seeking intensity* ($V$) tends to hurt the larger brand, an effect more pronounced the less similar the brands. However, for the Lattin–McAlister model, if the smaller brand seems very similar, is in effect 'subsumed' by the larger brand, the larger brand may benefit from increasing $V$.

*Adding a new unique feature* ($U_1$) is always beneficial to the brand receiving it and detrimental to the other brand.

*Increasing the value of a shared feature* ($S_{12}$) always benefits the larger brand.

## Comparative statics in a multi-brand market

Expanding the analysis by introducing additional brands allows an important step toward generality, in that it is possible to study how a change in two brands' similarity will affect a third, uninvolved brand, or how different share distributions affect parametric changes. While it is possible to explicitly calculate quantities analogous to (2)–(3) for any number of brands, the complexity of the resulting formulas rapidly outpaces ones ability to say anything concrete about them. Rather, we derive matrix-based solutions for equilibrium market shares and derivatives of interest ($(\partial X/\partial V)$, $(\partial X/\partial U_i)$ and $(\partial X/\partial S_{ij})$, where $X$ is the vector of market shares) in the general $n$-brand case. These precise solutions can in turn be incorporated into a simulation study to explore the model's robust properties.

## Model solution and derivatives

Given

$$A = [a_{ji}] \quad \text{for } 1 \le \{i, j\} \le n,$$

we have the standard Markov formulation, $X = XA$, where $X$ is the row-vector of steady-state probabilities (market shares). This gives way to the redundant system $0 = X(A\text{-}I)$, which can be made invertible by realizing that market shares must sum to one. This yields the system

$$XA^* = e_n = (0\ 0\ \cdots\ 0\ 1) \qquad (4)$$

where $A^*$ is just A-I with its last column replaced by ones. The solution to (4) is therefore

---

[2] With $\pi_1 - \pi_2$ and $f^2(\cdot)$ both non-negative, the sign of $(\partial x_1/\partial V)$ depends only on

$$S_{12}(1-VS_{12})^2 - (1-V^2S_{12})\pi_1\pi_2.$$

Since

$$1 - V^2S_{12} > 1 - VS_{12} > (1-VS_{12})^2,$$

$(\partial x_1/\partial V)$ will be non-positive whenever $S_{12} < \pi_1\pi_2$.

Since $1-V$, $\pi_2 - VS_{12}$ and $f^2(\cdot)$ are all non-negative, the sign of $(\partial x_1/\partial U_1)$ depends only on

$$(1-VS_{\cdot 2})(1-VS_{12}) - V\pi_2(\pi_1 - \pi_2).$$

With $\pi_1 > \pi_2$, this term is decreasing in $V$ and $S_{12}$, so it is minimized at $V=1$ and $S_{12} = \pi_2$, where its value is, after some algebra, $(\pi_1 - \pi_2)^2$, showing $(\partial x_1/\partial U_1)$ to be non-negative.

Similarly, since $V$, $1-V$, $\pi_1 - \pi_2$ and $f^2(\cdot)$ are all non-negative, the sign of $(\partial x_1/\partial S_{12})$ depends only on the term

$$(1-VS_{12})^2 - V\pi_1\pi_2,$$

which is decreasing in both $V$ and $S_{12}$. Replacing these with their extremal values, 1 and $\pi_2$, this term becomes $(1-\pi_2)^2 - \pi_1\pi_2$, which factors to $\pi_1(\pi_1 - \pi_2)$, showing $(\partial x_1/\partial S_{12})$ to be non-negative.

given by

$$X = e_n[A^*]^{-1} \tag{5}$$

To calculate the derivative of $[X]$ with respect to any model parameter, we compute $(\partial X/\partial \eta)$ for $\eta = V$, $U_i$, or $S_{ij}$. Since $XA^* = e_n$, we can take derivatives as follows:

$$0 = \frac{\partial e_n}{\partial \eta} = \frac{\partial [XA^*]}{\partial \eta} = \left[\frac{\partial X}{\partial \eta}\right] A^* + X\left[\frac{\partial A^*}{\partial \eta}\right]. \tag{6}$$

Solving for $(\partial X/\partial \eta)$,

$$\frac{\partial X}{\partial \eta} = -X\left[\frac{\partial A^*}{\partial \eta}\right][A^*]^{-1}$$

$$= -e_n[A^*]^{-1}\left[\frac{\partial A^*}{\partial \eta}\right][A^*]^{-1}. \tag{7}$$

We see that, to calculate $(\partial X/\partial \eta)$ for any model parameter of interest, the only information needed beyond the matrix $A^*$ is $(\partial A^*/\partial \eta)$, easily assembled from the elements $(\partial a_{ji}/\partial \eta)$. These individual derivative elements are calculated in the Appendix for each of the models here.

## Simulation

While there do exist statements about the models that can be proven in the $n$-brand case, there are others which, though not always true, are robust, and highlight general tendencies of the models in question. This latter type of statement is best attacked through simulation.

We will address four models here, the Givon and Modified Givon, the Lattin–McAlister and its parametric restriction, $S_{ij} = \pi_i \pi_j$ (Restricted L–M). For flexibility and ease of presentation, we use a three-brand choice set, with brand preferences chosen randomly and scaled to sum to one. For the unrestricted Lattin–McAlister model, this leaves the $S_{ij}$ terms undecided. The $\{S_{ij}\}$ are really free parameters, constrained only in that

$$0 \le S_{ij} \le \min\{\pi_i, \pi_j\}$$

and

$$\sum_{k \ne j} S_{kj} \le \pi_j,$$

and they will be chosen randomly subject to these constraints.

We address the following questions through simulation, each motivated by varying model parameters: What happens to the largest, middle and smallest preference brands overall as a result of variety-seeking $(X_i - \pi_i)$; As variety seeking intensifies, what can be expected to happen to each brand $(\partial X_i/\partial V)$; Should two brands derive more of their value through features they share, what will happen to these brands $(\partial X_i/\partial S_{ij})$ and $(\partial X_j/\partial S_{ij})$, and to an uninvolved brand $(\partial X_k/\partial S_{ij})$; What is the effect of a new, unique feature $(\partial X/\partial U_i)$. We simulate 5000 cases for each of the five models, tallying the proportion of time each of the following six statements, grouped to correspond to the above questions, is true (subscripts reflect preference order, $\pi_1 \ge \pi_2 \ge \pi_3$):

1. Overall, variety seeking is beneficial:
   a) $x_1 - \pi_1 > 0$,
   b) $x_2 - \pi_2 > 0$,
   c) $x_3 - \pi_3 > 0$.
2. Increasing variety seeking intensity is beneficial:
   a) $\partial x_1/\partial V > 0$,
   b) $\partial x_2/\partial V > 0$,
   c) $\partial x_3/\partial V > 0$.
3. Additional unique features benefit any brand and hurt all others:
   $\partial x_i/\partial U_j > 0$ iff $i = j$.
4. Should two brands begin to share more features, the larger is helped:
   a) $\partial x_1/\partial S_{12} > 0$,
   b) $\partial x_1/\partial S_{13} > 0$,
   c) $\partial x_2/\partial S_{23} > 0$.
5. Should two brands begin to share more features, the smaller is helped:
   a) $\partial x_2/\partial S_{12} > 0$,
   b) $\partial x_3/\partial S_{13} > 0$,
   c) $\partial x_3/\partial S_{23} > 0$.
6. Should two brands begin to share more features, the uninvolved brand is helped:
   a) $\partial x_3/\partial S_{12} > 0$,
   b) $\partial x_2/\partial S_{13} > 0$,
   c) $\partial x_1/\partial S_{23} > 0$.

The last three statements apply only to the Lattin–McAlister model and the Restricted Lattin–McAlister model. Because the completely general unrestricted Markov model and the Multinomial model have no parametric basis, they cannot be called on here, though later they will serve as benchmarks against which the fit of these other four models can be compared.

To test these statements, values were randomly selected for $V$ and, subject to appropriate summation constraints, for each of the individual features $\{U_i\}$ and $\{S_{ij}\}$; the precise solutions for $X_i$ and for each of the necessary derivatives (given in the Appendix) were then calculated. All analysis was run in the Matlab system, using a uniform random number generator based on the linear congruential method, as described in Forsythe et al. (1977). The results of the simulations are given in Table 1.

Post-analysis shows that the simulation covered the seven-dimensional parameter space admirably ($U_1$, $U_2$, $U_3$, $S_{12}$, $S_{13}$, $S_{23}$, $V$). For statement 3, which makes claims about the signs of nine derivatives at once, the listed numbers indicate that all nine always held. Only two of the statements were verified in all cases across all models; whether or not this indicates that statements 3 and 6 are logical consequences of these models is uncertain. Several of the other statements, 5 in particular, appear quite robust, though not unilaterally true and certainly unprovable.

Of the six statements, the one with the least decisive feel to it is the fourth, that, for the Lattin–McAlister model and its restricted variant, larger preference brands are helped by sharing more features. Recall that, for a two-brand choice set, this is always true (i.e. the larger brand always gains by sharing more features), and one might question what market conditions would predispose the larger preference of two brands to attempt to increase the perceptions of features it shares with a smaller competitor. Analysis of the simulation data reveals a clear trend: the larger the preference differential between two brands, the more likely the larger will be helped by an increase in shared features. However, in situations of mild share advantage in a multi-brand choice set, the larger of two brands will very likely *lose* share should it begin to appear more similar to a smaller competitor (who is certain to be hurt), while all other brands benefit. In any case, this share loss is less dramatic than that of the smaller brand and will, in many cases, represent a small proportion of the larger brand's overall preference.

Based on the results in Table 1, one might claim the following relative to the models at hand:

Table 1
Simulation results: Proportion of cases in which statement was correct ($n = 5000$ per cell)

| Statement | Model | | | |
|---|---|---|---|---|
| | Givon | Modified Givon | Restricted L–M | Lattin–McAlister |
| 1. Brands helped overall | | | | |
| a) $x_1 - \pi_1 > 0$ | 0 | 0 | 0 | 0.229 |
| b) $x_2 - \pi_2 > 0$ | 0.508 | 0.830 | 0.702 | 0.323 |
| c) $x_3 - \pi_3 > 0$ | 1 | 1 | 1 | 0.893 |
| 2. Brands benefit from increasing V | | | | |
| a) $(\partial x_1/\partial V) > 0$ | 0 | 0 | 0 | 0.195 |
| b) $(\partial x_2/\partial V) > 0$ | 0.508 | 0.773 | 0.752 | 0.360 |
| c) $(\partial x_3/\partial V) > 0$ | 1 | 1 | 1 | 0.896 |
| 3. Additional unique features help | | | | |
| $(\partial x_i/\partial U_j) > 0$ iff $i = j$ | 1 | 1 | 1 | 1 |
| 4. Larger helped by sharing | | | | |
| a) $(\partial x_1/\partial S_{12}) > 0$ | – | – | 0.678 | 0.770 |
| b) $(\partial x_1/\partial S_{13}) > 0$ | – | – | 0.904 | 0.925 |
| c) $(\partial x_2/\partial S_{23}) > 0$ | – | – | 0.522 | 0.502 |
| 5. Smaller helped by sharing | | | | |
| a) $(\partial x_2/\partial S_{12}) > 0$ | – | – | 0 | 0 |
| b) $(\partial x_3/\partial S_{13}) > 0$ | – | – | 0 | 0 |
| c) $(\partial x_3/\partial S_{23}) > 0$ | – | – | 0 | 0.0004 |
| 6. Uninvolved helped by sharing | | | | |
| a) $(\partial x_3/\partial S_{12}) > 0$ | – | – | 1 | 1 |
| b) $(\partial x_2/\partial S_{13}) > 0$ | – | – | 1 | 1 |
| c) $(\partial x_1/\partial S_{23}) > 0$ | – | – | 1 | 1 |

• The larger a brand's preference share, the more likely it is to be hurt overall by variety-seeking, and to be hurt by any increase in the level of variety-seeking intensity.

• When two brands begin to be perceived as sharing more features, the smaller one is certain to be hurt, while the larger one might actually benefit, particularly so when it offers more by way of unique features than its competitors. Regardless, all other brands will benefit.

We note that, with the exception of the unrestricted Lattin–McAlister model, each of the models makes unilateral statements about what tends to happen to the smallest and largest preference brands. This may indicate that the feature-based orientation of the Lattin–McAlister model allows a finer assessment of dynamics based on perceptions of interbrand similarity, a type of insight lost when all brands are considered to be composed solely of unique features.

In the following section, the model is fitted for each of the participants in a computer experiment. We then compare the degree of fit of the various parameterized models to one another and to two benchmarks, the completely general Markov model, which must fit best, and the Multinomial model, which assumes no choice dependence, and must fit most poorly.

## Fitting the model: An experiment

We designed a interactive computer experiment in which subjects choose among three hypothetical brands of shampoo, each described by its attributes. Pretests (80 subjects total) indicated that there was uncertainty about the effectiveness of different shampoos, and that shampoo is often chosen based on attributes touted in advertising or on the package itself, rather than on personal perceptions developed through use. Further, shampoo is a product class in which it is reasonable to anticipate both brand-loyal and variety-seeking segments.

Prior to the experiment, subjects were asked an assortment of questions about their purchasing patterns in several product categories, amongst them shampoo, as well as their attitude towards shampoo 'build-up'. Subjects prone to viewing build-up as a problem were assigned to one group ($n = 17$), while those not so predisposed were assigned to another ($n = 15$). To further encourage either variety-seeking or zero-order behavior, each group was presented with a different 'medical report', either extolling the virtues of switching among brands to prevent build-up (variety-seeking condition), or dismissing build-up as a concern (zero-order condition).

Subjects were given general instructions and shampoo brand descriptions, then asked to rate each brand on a scale of 1 to 100 to reflect their relative preferences (this allows a between-group test of preference homogeneity across the three brands; it cannot be rejected, $p > 0.10$). Finally, subjects made a series of choices among the three brands. After each choice, to help prevent automatic behavior, there was a pause, the brands were re-ordered and the screen color was changed.

## Fitting and tests

Each of the 32 subjects generated a string of 40 purchases. It is possible to use these purchase histories to fit the six models in question. With three brands, The unrestricted Markov and Lattin–McAlister models each have six estimated parameters, the Restricted Lattin–McAlister, Givon and Modified Givon models each have three, while the Multinomial has two.

As a test of the variety-seeking manipulation and the validity of the $V$ parameterization, we fit the Lattin–McAlister, Restricted Lattin–McAlister, Givon and Modified Givon models and compare the average $V$ between the two groups; note that these parameters should not be compared across models, as they measure different things. For the Lattin–McAlister model, the average $V$ for the variety-seeking and zero-order groups were, respectively, 0.68 and 0.27 ($p < 0.01$); for the Restricted Lattin–McAlister model, average $V$ values were 0.62 and 0.26 ($p < 0.01$); for the Givon model, average $V$-values were 0.36 and 0.06 ($p < 0.001$); for the Modified Givon model, average $V$-values were 0.29 and 0.05 ($p < 0.001$). This offers evidence both of the effectiveness of the variety-seeking manipulation and the ability of each of the four models to highlight its presence. Although an unrestricted Markov formulation must fit at least as well, it affords no insight into the degree of variety-seeking intensity.

### Tests

As discussed in Hauser (1978), Horowitz (1983) and Rust and Schmittlein (1985), the relative performance of differing stochastic model formulations can be easily compared using the likelihood ratio test, but only when they are nested. When this is the case, twice the likelihood ratio is asymptotically distributed Chi-square, with degrees of freedom equal to the difference in estimated parameters of the two models. Since all of the models are nested within the unrestricted Markov (i.e., are special-case parametric restrictions of it), and have the Multinomial further nested within them (i.e., setting any of their $V$-parameters to zero), it is possible to use this likelihood ratio test to compare models in this fashion.

When the models are non-nested, no such simple methodology is available, though a variety of methods exist to partially clarify the situation. Hauser (1978) develops an information-theoretic measure ($U^2$), an intuitively appealing generalization of the standard likelihood ratio index ($\rho^2$), which does not account for the number of estimated parameter differences in the models; Horowitz (1983) offers a variant of $\rho^2$ which heavily penalizes models that are not parsimonious; Akaike (1974) offers an information-theoretic criterion, different from Hauser's, which penalizes lack of parsimony even more strongly than that of Horowitz; Rust and Schmittlein (1985) thoroughly review applicable Bayesian techniques. Each of the available methods has its shortcomings. Since Hauser's statistic is equivalent to the likelihood ratio index, and because it is unclear how to test for *significant* differences in non-nested model fits, we will use $U^2$ to compare pairs of non-nested models. For the models that can be nested, we will use the asymptotically Chi-square distributed likelihood ratio test, which does account for degree-of-freedom differences. For a technical discussion of the discriminatory ability of these and related statistics, see Horowitz (1983).

In Table 2 are compiled, for each pair of models, the number of subjects for which the likelihood ratio test shows a significant difference for both the variety-seeking and zero-order groups at the 0.05 level. Since the unrestricted Markov model is guaranteed to fit at least as well as any other first-order model, the first line (for both the variety-seeking and zero-order groups) reports the number of subjects for which the unrestricted Markov model fit *significantly better* than the listed models. Similarly, since the Multinomial model fits no better than any first-order model, the second line reports the number of subjects for which the Multinomial model fit *significantly worse* than the listed models. The degrees of freedom on the Chi-square statistic for these comparisons were, respectively, three and one; tests were carried out at the 0.05 level. The Lattin–McAlister model is excluded here, as it has the same number of estimated parameters, when there are three brands, as the unrestricted Markov model, rendering the likelihood ratio test inappropriate; for larger number of brands, such a comparison would be possible.

As anticipated, the variety-seeking group affords the best setting for distinguishing between the various models, while each model fits about

Table 2
Comparing five models: Number of subjects for which models can be distinguished by a likelihood ratio test ($p < 0.05$)

| | Unrestricted Markov | Restricted L–M | Modified Givon | Givon | Multinomial |
|---|---|---|---|---|---|
| *Variety-seeking group:* | | | | | |
| Number of times ($n = 17$) Unrestricted Markov fit better | – | 7 | 10 | 12 | 13 |
| Number of times ($n = 17$) Multinomial fit worse | 13 | 11 | 9 | 7 | – |
| *Zero-order group:* | | | | | |
| Number of times ($n = 15$) Unrestricted Markov fit better | – | 1 | 1 | 2 | 2 |
| Number of times ($n = 15$) Multinomial fit worse | 2 | 0 | 0 | 0 | – |

equally well within the zero-order group; we note that the qualitative results drawn from these tables would be equally convincing at $\alpha = 0.01$ or $\alpha = 0.10$ as they are at $\alpha = 0.05$.

The results of Table 2 illustrate a somewhat unexpected hierarchy of sorts between the models. Although, within the zero-order group, the only model which performed a bit better than the others was the unrestricted Markov, the results from the variety-seeking group show a marked dominance as one proceeds across the table. By far the poorest showing came from the Multinomial model; adding just one more parameter, as do the Givon, Modified Givon and Restricted Lattin–McAlister models, a significant improvement in fit is observed for at least half the subjects.

The results of applying Hauser's $U^2$ to compare non-nested pairs of models are compiled in Table 3. For any model pair, the number of subjects is listed, within the variety-seeking group, for which one model provided a higher $U^2$-value than the other. As there were four subjects for which all the models fit equally well (fitted $V$ near zero indicates these subjects were apparently zero-order), we restrict the sample size to thirteen. It is difficult to interpret these values in terms of significance; we refer only to the variety-seeking group, as differences within the zero-order group were, in any case, extremely small.

The data of Table 3 offer a tentative hierarchy of the models. It is not surprising that the Lattin–McAlister model fares best here; it is the least parsimonious of the models. However, the Restricted Lattin–McAlister model performs quite well on its own, edging out the Modified Givon model for slightly more than half the subjects, and the Givon model for almost all of them. In any row, the upward trend as one reads across the table lends some credence to the models being ordered here in terms of performance. The chief advantage of the Lattin–McAlister formulation, which appears to fit well even when its similarity parameters are restricted, lies in its ability to clarify the similarity structure of the market. Whether this more complete fit compensates for the loss of parsimony it entails falls under the discretion of the decision-maker.

## Summary and conclusions

We have focused on a variety of first-order Markov models of variety seeking, catalogueing their parametric properties and comparing their relative ability to describe variety-seeking consumers' purchase sequences. Although a modest performance hierarchy can be set up between the models, we have made no attempt to demonstrate any type of optimality relative to the wide variety of possible variant models that can be formulated; certainly, the mechanisms for encoding variety-seeking behavior into stochastic models are in no way exhausted. Further, we have not considered the question of likely competitive response.

Still, some support can be found for the notion that a feature-based approach, like that offered by the Lattin–McAlister model, may be more desirable than one based solely on brand preference. For our sample of variety-seeking consumers, it is not surprising that the Multinomial model, ignoring as it does both feature-sharing and variety-seeking in general, fits least well. The Givon model, while it does account for variety-seeking, does so in a way that ignores relative brand preferences once the consumer decides to seek variety. The Modified Givon model partially remedies this, and consequently fits a bit better.

Table 3
Number of subjects in variety-seeking group ($n = 13$) for which model $i$ yields a higher $U^2$ than model $j$

| Model $i$ | Model $j$ | | | |
|---|---|---|---|---|
| | Lattin–McAlister | Restricted L–M | Modified Givon | Givon |
| Lattin–McAlister | – | 11 | 11 | 12 |
| Restricted Lattin–McAlister | 2 | – | 7 | 12 |
| Modified Givon | 2 | 6 | – | 10 |
| Givon | 1 | 1 | 3 | – |

Yet, even with the same number of estimated parameters, the Restricted Lattin–McAlister model fits arguably better than any of these, by making a hypothesis about how likely brands are, on average, to share features, and how this feature-sharing affects the conditional probability of brand-switching. While these results can hardly be called conclusive, they do offer moderate evidence that a feature-based approach can work better than one formulated on brand preference alone.

## Appendix

Recall that the transition probabilities, $a_{ji}$, are given by:
Lattin–McAlister:

$$a_{ji} = \frac{\pi_i - VS_{ij}}{1 - VS_{\cdot j}},$$

Givon:

$$a_{ji} = \begin{cases} (1-V)\pi_i & \text{for } i = j, \\ (1-V)\pi_i + \dfrac{V}{n-1} & \text{otherwise,} \end{cases}$$

Modified Givon:

$$a_{ji} = \begin{cases} (1-V)\pi_i & \text{for } i = j, \\ (1-V)\pi_i + V\dfrac{\pi_i}{1-\pi_j} & \text{otherwise.} \end{cases}$$

*Changing V:* We merely need the individual entries $(\partial a_{ji}/\partial V)$:
Lattin–McAlister:

$$\frac{\partial a_{ji}}{\partial V} = \frac{S_{\cdot j}\pi_i - S_{ij}}{(1 - VS_{\cdot j})^2},$$

Givon:

$$\frac{\partial a_{ji}}{\partial V} = \begin{cases} -\pi_i & \text{for } i = j, \\ \dfrac{-\pi_i + 1}{n-1} & \text{otherwise,} \end{cases}$$

Modified Givon:

$$\frac{\partial a_{ji}}{\partial V} = \begin{cases} -\pi_i & \text{for } i = j, \\ \dfrac{\pi_i \pi_j}{1-\pi_j} & \text{otherwise.} \end{cases}$$

*Changing $U_l$:* Without loss of generality, we focus on brand $B_1$; we need to calculate $(\partial a_{ji}/\partial U_1)$. Increasing the value of $U_1$ requires that each of the model's other feature parameters (i.e. $U$'s and $S$'s) be rescaled so that the ratio of its old value to its new value is constant. Recalling that the $\pi_i$ are constrained to sum to 1, we have that, for $\eta$ other than $U_1$,

$$\frac{\partial \eta}{\partial U_1} = -\eta,$$

while $(\partial U_1/\partial U_1) = 1$. This allows us to compute the required matrix elements:
Lattin–McAlister:

$$\frac{\partial a_{ji}}{\partial U_1} = \begin{cases} \dfrac{-a_{ji}}{(1-U_1)(1-VS_{\cdot j})} & \text{for } i > 1, \quad j > 1, \\ \dfrac{1-a_{ji}}{(1-U_1)(1-VS_{\cdot j})} & \text{for } i = 1, \quad j > 1, \\ \dfrac{-a_{ji}(1-V)}{(1-U_1)(1-VS_{\cdot j})} & \text{for } i > 1, \quad j = 1, \\ \dfrac{(1-a_{ji})(1-V)}{(1-U_1)(1-VS_{\cdot j})} & \text{for } i = 1, \quad j = 1, \end{cases}$$

Givon:

$$\frac{\partial a_{ji}}{\partial U_1} = \left\{ \begin{array}{ll} 1-V & \text{for } i = 1, \\ -(1-V)\pi_i & \text{otherwise,} \end{array} \right\}$$

Modified Givon:

$$\frac{\partial a_{ji}}{\partial U_1} = \begin{cases} 1-V \\ \quad \text{for } i = 1, \quad j = 1, \\ -(1-V)\pi_i \\ \quad \text{for } i > 1, \quad j > 1, \; i = j, \\ 1 - V + V\dfrac{(1-\pi_j-\pi_i\pi_j)}{(1-\pi_j)^2} \\ \quad \text{for } i = 1, \quad j > 1, \\ -(1-V)\pi_i + V\dfrac{\pi_i\pi_j}{(1-\pi_j)^2} \\ \quad \text{for } i > 1, \quad j = 1, \\ -(1-V)\pi_i - V\dfrac{\pi_i}{(1-\pi_j)^2} \\ \quad \text{otherwise.} \end{cases}$$

*Changing $S_{12}$ with $\pi$'s held constant:*

Without loss of generality, we choose two brands, $B_1$ and $B_2$, with which to carry out our analysis. Since only $S_{12}$ is changing, none of the $\pi_j$ change for any $j$, and of the $U_j$, only $U_1$ and $U_2$ change their value in that they each decrease by the amount by which $S_{12}$ increases; so, $a_{ji}$ is entirely independent of $S_{12}$ for $i$ and $j$ both greater than 2:

Lattin–McAlister:

$$\frac{\partial a_{ji}}{\partial S_{12}} = \begin{cases} 0 & \text{for } j > 2, \\ \dfrac{V(a_{ji} - 1)}{(1 - VS._j)} & \text{for } (i, j) = (1, 2) \\ & \quad \text{or } (2, 1), \\ \dfrac{Va_{ji}}{(1 - VS._j)} & \text{otherwise.} \end{cases}$$

## References


Akaike, H. (1974), "A new look at the statistical model identification", *IEEE Transactions on Automatic Control* 19, 716–723.

Bass, F.M., Pessemier, E.A., and Lehmann, D.R. (1972), "An experimental study of relationships between attitudes, brand preference, and choice", *Behavioral Science* 17, 532–541.

Forsythe, G., Malcolm, M., and Moler, C. (1977), *Computer Methods for Mathematical Computations,* Prentice-Hall, Englewood Cliffs, NJ.

Givon, M. (1984), "Variety seeking through brand switching", *Marketing Science* 3/1, 1–22.

Hauser, J.R. (1978), "Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach", *Operations Research* 26, 406–421.

Horowitz, J.L. (1983), "Statistical comparison of non-nested probabilistic discrete choice models", *Transportation Science* 17/3, 319–350.

Hutchinson, J.W. (1986), "Discrete attribute models of brand switching", *Marketing Science* 5/4, 350–371.

Jeuland, A.P. (1979), "Brand choice inertia as one aspect of the notion of brand loyalty", *Management Science* 25/7, 671–682.

Jones, J.M. (1973), "A composite heterogeneous model for brand choice behavior", *Management Science* 19/5, 499–509.

Lattin, J.M., and McAlister, L. (1985), "Using a variety seeking model to identify substitute and complementary relationships among competing products", *Journal of Marketing Research,* August.

McAlister, L., and Pessemier, E.A. (1982), "Variety seeking behavior: An interdisciplinary review", *Journal of Consumer Research* 9, 311–322.

Morrison, D.G. (1966), "Testing brand switching models", *Journal of Marketing Research* 3, 401–409.

Rust, R.T., and Schmittlein, D.C. (1985), "A Bayesian cross-validated likelihood method for comparing alternative specifications of quantitative models", *Marketing Science* 4/1, 20–40.