JAGDISH N. SHETH*

With factor analysis as a method of estimating parameters, an empirical model of measuring brand loyalty for individual consumers based on frequency and pattern of purchases is presented. Despite some limitations, the method seems superior to stochastic models for generating robust measures at the individual level.

A Factor Analytical Model of Brand Loyalty

INTRODUCTION

A variety of approaches have been introduced recently for understanding buyer behavior and particularly the phenomenon of brand loyalty and brand switching [26]. Operations research techniques have been extensively applied in this direction with the availability, for academic research, of panel data, notably from the *Chicago Tribune* and the Marketing Research Corporation of America. Several stochastic models of consumer behavior, in which some measure of time as the independent variable and some measure of brand loyalty as the dependent variable, have been extended and tested using panel data. For example, Kuehn [15, 16] has developed a linear learning model that states that probability of buying a brand at time t depends on the sequence of all past purchases before time t. Frank [8], however, finds evidence suggesting that probability of buying for each buyer remains constant over time and therefore independent of purchase history. Most researchers have argued that dynamics of brand loyalty is a first-order Markov process [1, 3, 4, 7, 11, 12, 17–19, 21, 22, 27, 34]. They feel that purchase probability, as a measure of brand loyalty, at time t is dependent only on the immediate past purchase at $t-1$, and any earlier history is irrelevant for prediction.

Compared with traditional measures of market share and proportion of repeat buyers, the diffusion of stochastic models of consumer behavior has tremendously enriched our measures of brand loyalty. However, the panel data used in these models are gathered for monitoring market behavior and *not* for testing any specific stochastic model with its set of assumptions.

*Jagdish N. Sheth is assistant professor of business, Columbia University.

Second, consumer behavior is complex enough that no single theoretical model is appropriate to all products and consumers. There is, therefore, a need for an empirical model similar to the curve-fitting technique that will *derive* measures of brand loyalty for each situation.

Yet another and important consideration is that existing stochastic models have so far provided measures of aggregate brand loyalty for the whole market. What we need is a model that will, besides aggregate brand loyalty measures, provide measures of each consumer's brand loyalty.

Based on rank reduction procedures of matrix algebra, a factor analytic model of brand loyalty is proposed and tested in this article. The model is empirical in that it generates measures of brand loyalty from a set of panel data for which there need not be any a priori hypothesis on the form or shape of the brand loyalty curve over time. However, if some hypothesis does exist, the model can test whether the theory holds in that particular situation. More interesting and valuable is the model's capability to generate brand loyalty scores separately for each person in the sample. The distribution of such scores then becomes the starting point for further analysis for market segmentation.

DESCRIPTION OF THE MODEL

Underlying consumer behavior analysis, we seem to use implicitly the S-O-R paradigm proposed by Woodworth [35] in psychology. The paradigm states that R (response function) is the manifested behavior or response that is dependent on both the environmental stimulus (S) and the state of the organism (O). The manifested response thus depends on (a) stimulus condition and (b) the person. Needed, therefore, is a

Journal of Marketing Research, Vol. V (November 1968), 395–404

formulation that will resolve a buyer's manifested behavior between the environmental effects and the individual buyer himself. Such resolution will separate the environmental and individual influences (parameters) at those points in time when the researcher obtains measures of purchase behavior.

Estimation by Factor Analysis

Tucker [31] provided a useful approach to individual parameter estimations of several linear and nonlinear functions. Take a function:

$$(1) \qquad y = f(a, b, c, \ldots, x),$$

where y represents the dependent variable, x the independent variable, and $a, b, c, \cdots$ are constants of the curve. In buying, for example, y may represent the probability of a brand's purchase and x may be the time periods or trials ($x = 1, 2, 3, \cdots, n$); $a, b, c, \cdots$ etc. are the parameters of the functional relation. The buyer's function can be generally represented as

$$(2) \qquad y_i = f(a_i, b_i, c_i, \ldots, x).$$

If j is any particular point of this function with coordinates x_j and y_{ji}, we have

$$(3) \qquad y_{ji} = f(a_i, b_i, c_i, \ldots, x_j).$$

Thus a given observation y_{ji} is a function of the values of individual parameters and the value of the independent variable, namely x_j. Several functions, both linear and nonlinear, may be transformed to produce

$$(4) \qquad y_{ji} = \sum_{m=1}^{r} f_m(x_j)F_m(G_i).$$

The $f_m(x_j)$ are a number of functions of the independent variable x_j. Similarly, the $F_m(G_i)$ are functions of the parameters $G_i = a_i, b_i, c_i, \cdots$ etc. Thus an observation y_{ji}, which results from the interaction of the person and the environment, is a sum of the products of functions of the individual parameters and corresponding functions of the independent variable. The problem now is to find some technique to provide estimates of $f_m(x_j)$ and $F_m(G_i)$. Equation 4 is analogous to the basic linear postulate of factor analysis. If we define

$$f_m(x_j) \equiv a_{jm} \quad \text{and} \quad F_m(G_i) \equiv s_{mi},$$

then the basic factor analytic postulate is readily obtained:

$$(5) \qquad y_{ji} = \sum_{m=1}^{r} a_{jm}s_{mi}.$$

The transformation process analyzes a functional relation into m linear dimensions very similar to resolution of complex wave forms, from sound or heat vibration, into simpler component sine-wave curves by Fourier analysis.

For example, suppose our theory dictated that, for a finite length of time, a brand's purchase probability is independent of the history of purchases. Then for a point in time j the probability of a consumer i purchasing is y_{ji}. Since it is independent of time, we have a functional relation,

$$y_{ji} = a_i,$$

where a_i is some level of probability that can range from zero and one and is a specific parameter of consumer i.

If we define

$$a_{j1} = 1$$

and

$$s_{1i} = a_i,$$

then

$$y_{ji} = a_{j1}s_{1i}.$$

There is only one linear dimension. Furthermore, if we plot the a_{j1} values on a graph where the x-axis is consecutive trials or time periods and the y-axis represents a_{j1} values for each $j (j = 1, 2, 3, \cdots, n)$, we should expect a horizontal straight line since a_{j1} is a constant by definition.

If, however, our theory dictates that the purchase probability of the brand is some linearly increasing function of time periods or trials, then for a given consumer i at point j, we have

$$y_{ji} = a_i + b_ix_j.$$

This function can be transformed into Equation 5, if we define

$$a_{j1} = 1 \qquad s_{1i} = a_i$$
$$a_{j2} = x_j \qquad s_{2i} = b_i,$$

and write

$$y_{ji} = a_{j1}s_{1i} + a_{j2}s_{2i} = \sum_{m=1}^{2} a_{jm}s_{mi}.$$

The linearly increasing function is resolved into two dimensions ($m = 2$). If we again plot a_{j1} values for $j = 1, 2, 3, \cdots n$, it will be a straight line since a_{j1} is a constant. However, if we plot a_{j2} values for $j = 1, 2, 3, \cdots n$, we will have a positively inclined straight line since values of x_j are ordered.

Finally, suppose our theory dictates that the purchase probability is some exponential function of time. Suppose the function looks as follows for a given consumer i at a point j:

$$y_{ji} = e^{b_i + x_j}.$$

We can rewrite this as

$$y_{ji} = (e^{b_i})(e^{x_j}),$$

and define

$$a_{j1} = e^{x_j} \qquad s_{1i} = e^{b_i},$$

so that

$$y_{ji} = a_{j1}s_{1i} .$$

This exponential function can be resolved in one dimension. Again, if we plot the values of a_{j1} for $j = 1, 2, 3, \cdots n$, we should expect values that are powers of the natural log base e.

The a_{jm} and s_{mi} coefficients obtained by linear transformation are parameters of the functional relation. These parameters are more difficult to interpret than those derived by fitting a least-squares curve since the common meanings of intercept and slope are lost in the process of transformation. However, they serve the same function of obtaining the best fit to the data. We could, therefore, use them to predict purchase behavior in the future.

The a_{jm} coefficients are aggregate parameters for each of the j time periods or trials. In the S-O-R paradigm discussed earlier, the a_{jm} coefficients summarize the influence of S or the environment. As such, they remain constant across a sample of consumers. Later, we will show that they determine aggregate purchase behavior and hence tell whether it is a monotonically increasing, decreasing, or constant function of time. When we plot each dimension m using a series of a_{jm}, a curve results which we will call a "reference curve" [31]. There will be as many curves as there are dimensions of a functional relation. A family of reference curves then summarizes the aggregate complexity of functions. In fact, the function's complexity determines the number of dimensions. To put it differently, a function can be unidimensional or multidimensional. The number of reference curves will tell us a function's dimensionality.

So far we have said nothing about the s_{mi} values. Note that each consumer will have as many s_{mi} values as there are linear dimensions m of the function. In the first and third types of functions shown earlier, there was only one dimension ($m = 1$); therefore, the consumer i had only one value. In the second type of function, however, there were two dimensions ($m = 2$), and the consumer i had two values. These values are the derived parameters of the functional relation and serve the same function as the primitive parameters obtained by the usual estimation procedures. Thus in the second type of functional relation, i.e., $y_{ji} = a_i + b_ix_j$, the first value s_{1i} will indicate the level of brand loyalty for the consumer i, and the second value s_{2i} will indicate both the extent and the direction (whether positive or negative) of change as a function of time.

The resolution of a function into linear dimensions is based on matrix algebra theorems that create a least-squares (minimization of error) estimation of the derived parameters. The procedure is described in the appendix. However, it is useful here to compare the method with the curve-fitting technique. First, in curve fitting, the parameters estimated are for the aggregate sample and not for each person in the sample. The parameters derived here are for each person i separately. Second, we simultaneously obtain separate estimates of parameters for all i buyers that are stored in a matrix as column vectors. In the first type of function in our example, i.e., $y = c$, a single dimension such as s_{1i} summarizes all possible values from zero to one. We may, for instance, simultaneously estimate and store $S_{11} = .4$ and $S_{12} = .8$ for two consumers having the same functional relation but different levels. Such estimation and summary of parameters of each of the consumers is even more useful when there are two dimensions as in the second type of function ($y_{ji} = a_i + b_ix_j$) in our example. With only two dimensions, we can aggregate all persons who have an infinite variety of linear function *shapes* (determined by the value of a_i and b_i). The extreme cases in this function would be:

1. $y_{ji} = a_i , b_i = 0,$
2. $y_{ji} = b_i , a_i = 0,$
3. $y_{ji} = a_i + b_ix_j , a_i \neq 0, b_i \neq 0.$

The first consumer has a positive value for s_{1i} and a zero value for s_{2i}; he will therefore have a constant purchase probability during our study's time interval. The second consumer has a zero value for s_{1i} and a positive value for s_{2i}; he will therefore start at the zero level, and his probability will increase upward linearly. The third consumer has a positive value for both s_{1i} and s_{2i}; he will initially start at a nonzero probability level that will increase linearly with time. We could not have aggregated a sample of consumers and still maintained various shapes for each of the three consumers in ordinary curve fitting.

DEFINITION OF BRAND LOYALTY

Our interest here is to use the factor analytic model described in the appendix to estimate each consumer's extent of brand loyalty, which is an index of some observed phenomenon as are "social class" and "probability of an event." Like any index, brand loyalty is likely to have several indicators all manifested by the consumer. For example, extent of purchase at a point in time or the frequency of purchase of a brand during a certain time interval are two possible indicators of brand loyalty. However, the most commonly used data are the number of times the consumer buys and the relative frequency of buying a given brand. Suppose a consumer purchased the brand we are interested in three in five times during a time interval, and the other two times he bought another brand. We may show this as a sequence 01011, in which 1 represents purchase of the brand we are interested in and 0 represents purchase of any other brand. Using the relative frequency of three in five purchases, we may state that his brand loyalty is .60. This is tantamount to his purchase probability if we assume no sampling problems.

However, the 01011 sequence provides additional information on the purchase *pattern*, which is crucial in any time-dependent process. The pattern is ignored when we use only the probability measure as described earlier because three purchases of the brand, however arranged, will give the same answer. It would appear, based on empirical evidence, that pattern will be important when new products are introduced and when the consumer is completely unaware of a product class. Even among mature, well-established product classes such as detergents and coffee, the pattern of buying a brand becomes important because constantly changing market conditions will disrupt any tendency toward stable purchase behavior of the consumer. Occasionally, however, we may find a period of stagnant marketing activity for a well-established product class during which brand loyalty may be simply a function of relative frequency and pattern may not be important.

Brand loyalty then is a function of a brand's relative frequency of purchase in time-independent situations, and it is a function of relative frequency and purchase pattern for a brand in time-dependent situations. The factor analytic model of brand loyalty handles both situations.

DESCRIPTION OF FACTOR ANALYTIC MODEL OF BRAND LOYALTY

To estimate brand-loyalty parameters as a function of time for each consumer, we need a sample of consumers who all manifest the purchase behavior. Suppose there are five consumers and three trials; record the purchase of a given brand as one and the purchase of any other brand as zero. Then put the raw data into a matrix Y with three rows for the trials and five columns for the consumers:

		Consumer (i)				
		1	*2*	*3*	*4*	*5*
	1	0	1	0	0	1
Trial (j)	2	1	1	0	1	1
	3	1	0	1	1	1

Each cell y_{ji} contains data on purchase behavior that is caused by interaction of consumer i with the environment at trial j. Each row j is a vector containing information on the outcomes at trial j among a sample of consumers. It summarizes the aggregate sample purchase activity classified as dichotomous events. For example, of these five consumers, two bought the brand on the first trial compared with four who bought the brand on the second and third trials. The a_{jm} parameters in Equation 5 are derived from these row vectors.

Each column i is a vector that contains information on the purchase pattern of each consumer i. For example, the first consumer began to buy the brand on the second trial and repeated the purchase on the third trial. The fifth consumer, however, bought the brand on all three trials. The s_{mi} parameters in Equation 5 are derived from the column vectors.

To estimate the parameters, the factor analytic model (see the appendix) transforms the n row and the N column vectors as points that are projected on a space with as many dimensions as there are reference curves of the underlying functional relation. This projection process also obtains a least-squares solution that minimizes error in the total system.

We may state Equation 5 in matrix notation to summarize several functions of a sample of consumers. Then

$$Y_{n\times N} = A_{n\times r}S_{r\times N}, \qquad (6)$$

where Y is a completely filled rectangular matrix ($n < N$) containing y_{ji} cells, A is an $n \times r$ matrix containing a_{jm} cells, and S is an $r \times N$ matrix containing s_{mi} cells. Each column of the A matrix is a dimension of the function and therefore a reference curve as already discussed. Each column of the S matrix contains parameters for each i consumer. Both the sign and size of a set of s_{mi} parameters for consumer i determine his functional relation.

When several consumers, each having a functional relation of the form of Equation 2, are aggregated into a data matrix Y, the first reference curve a_{j1} ($j = 1, 2, 3, \cdots, n$) approximates the average curve, which is the plot of average values at j time points or trials. If the function has only one dimension, as did the first function in our earlier example, the first reference curve is identical to the mean curve. If a function is complex and requires more than one dimension, the second, third, fourth, etc. reference curves represent deviations from the first reference curve.[1] As previously mentioned, a functional relation may be multidimensional and therefore may have a family of reference curves.

---

[1] This stems from the process of aggregation in which each function is set as $y_i = (\bar{a} + \alpha, \bar{b}. + \beta, \bar{c} + \gamma, \ldots x)$, i.e., expressed as deviation from the means of the constants. This also enables us to establish criteria for the legitimacy of aggregating persons having a given functional relation. As a formal criterion, all individual functions, having second and higher order partial derivatives for constants that become zero when the function is expanded by Taylor's series, can be aggregated without the aggregate function's form becoming different from that of the persons. The same functions can also be exactly transformed into the linear postulate. Other functions, having second or higher order partial derivatives that remain nonzero, can only be *approximated* by a finite number of reference curves [6, 28, 30].

Standardization of Different Sample Sizes

The first step in resolving a data matrix Y consisting of n trials and N consumers is to post-multiply it by its transpose Y' to obtain a square symmetrical matrix YY'. The YY' is a cross-products matrix in which the diagonal values are the sums of squares and the off-diagonal values are the sums of cross products of values in the original data matrix. In our example of three trials and five consumers, if we multiply the Y matrix by its transpose Y', we obtain YY' which is:

		Trials		
		1	*2*	*3*
	1	3	2	2
Trials	2	2	3	2
	3	2	2	3

As can be seen, when the cell values of the data matrix Y are dichotomies, the cross-products matrix YY' is a square symmetric contingency table. The diagonal values indicate the frequency of consumers who bought the brand at each of the j trials. The off-diagonal values give us the switching pattern in the sample. For example, three of five consumers bought the brand on the first trial, and of these three consumers, two continued to buy the brand, but one switched to another brand. Similarly, three of five consumers bought the brand at the second trial, and two of them continued to buy the brand, but one switched at the third trial.

In estimating the S matrix from the data matrix Y (actually YY' matrix), the parameter coefficients are rescaled to maintain the orthogonal property of the S matrix, i.e., $SS' = I$. The orthogonal property is essential to factor analytic procedures. Since the S matrix is $r \times N$, the coefficients are a function of the number of consumers in the sample. Then even if two sets of data differ only by sample size, the resulting coefficients are not comparable. Therefore, it is imperative that data be standardized to remove the effects of sample size. This can be done either before or after the analysis; a description of standardization follows.

To remove sample-size bias, it is necessary to express the cross-products values not in terms of the absolute but rather the relative joint frequencies. Therefore, we may state $P_{jk} = N_{jk}/N$ where N is the total number of buyers. The result will be a cross-products matrix containing values only between zero and one. Dividing the relative frequencies P_{jk} by the standard deviation $\sqrt{N_j N_{k/N}}$ [2] yields a set of proportionate values that may be designated as

$$Y'_{jk} = P_{jk}/\sqrt{P_j P_k} = N_{jk}/\sqrt{N_j N_k}.$$

This is equivalent to pre- and post-multiplying the cross-products matrix YY' by a diagonal matrix $D^{-1/2}$ having the elements $1/\sqrt{N_j}$. We thus arrive at a standardized square symmetric matrix, for example R.

$$R = D^{-1/2}\, YY' D^{-1/2}. \tag{8}$$

The matrix R is positive semidefinite. Being symmetrical, it has grammian properties. The standardization results in placing ones in the diagonal, making it suitable for direct principal components analysis.

The analysis yields a set of reference curves. The first reference curve will act as a mean curve and therefore approximate the relative frequency of purchases of the brand at each trial. However, the values will differ because of the second step in the standardization. A peculiarity exists because the second and other reference curves usually contain negative values, making isomorphism with probability notions untenable. With appropriate rotation of the principal axes (reference curves) we can easily avoid negative coefficients. However, since second and other reference curves act as correction terms for the first reference curve, it may be interesting to work with negative coefficients.

In the appendix, several procedures (size of the latent roots, runs test on reference curves, and calculation of the sum of the squares of first differences among consecutive coefficients a_{jm} called Σd^2) can be used to determine a finite number of significant reference curves.

DEVELOPMENT OF BRAND LOYALTY OF RICE

This section describes a specific application of the model. I gathered the data [24] to test central concepts of the Howard-Sheth theory of buyer behavior [14]. A main interest was to observe the development of brand loyalty from its beginning. The basic hypothesis was that development of brand loyalty over time is a process of learning by repeated purchases of a brand. Since for many consumer grocery and personal care items, buyers establish brand preferences early in life [9, 10, 20, 33], existing commercial panel data are not suitable. Therefore, a panel was formed of foreign students as soon as they came to the United States. Foreign students are unfamiliar with U.S. brands in several product categories and learn to form attitudes about them once they reach this country. Several measures of brand generalization and personal influence were obtained to prevent learning brand loyalty by sources other than repeated experiences [25]. The factor analytic model of brand loyalty was applied to several data matrices. We will restrict our discussion here to the development of brand loyalty of rice.

[2] The division of P_{jk} by standard deviation becomes relevant when we deal with multiple brands simultaneously as separate states and have, in essence, a manifold contingency table [2]. However, it is not essential when we deal with only two state systems as in binary data. It is incorporated here because it enables us to use existing computer programs.

Analysis of Data

The first step in the analysis is to form the data matrix Y with y_{ji} elements representing the purchase behavior of the ith buyer at the jth trial. Since our interest is not focused on a particular brand but rather on the loyalty of the buyer toward any brand, we may take the brand with the maximum purchase frequency as the preferred brand and code it as one and all other brands as zero [24, pp. 125–30].

The data matrix analyzed refers to the purchase of rice by several panel members. A sequence of seven trials was chosen for analysis, the sample size was 14. Thus, data matrix Y has seven rows and fourteen columns. Using the procedures described earlier, the data matrix was standardized to obtain the R matrix.

The R matrix is presented in Table 1. The standardization has resulted in placing ones in the diagonal and proportionate values in the off-diagonal elements. Table 2 provides the first three reference curves. The first reference curve explains a little more than 85 percent of the total variance and seems to be the only dominant curve. The runs test performed on all three reference curves shows that only the first curve is significant. The corresponding characteristic (latent) roots also support this: the first root is very large (5.97) and the second and third roots are much smaller (.62 and .15). Finally, the Σd^2 calculations suggest that only the first curve is very smooth. Therefore, only one significant reference curve is present in this data.

Table 1

STANDARDIZED PRODUCT-SUM MATRIX FOR RICE BRAND LOYALTY

(N = 14, n = 7)

	1	*2*	*3*	*4*	*5*	*6*	*7*
1	1	.707	.650	.707	.554	.500	.597
2	.707	1	.940	.877	.877	.823	.886
3	.650	.940	1	.940	.940	.886	.944
4	.707	.877	.940	1	.877	.877	.886
5	.554	.877	.940	.877	1	.940	.886
6	.500	.823	.886	.877	.940	1	.944
7	.597	.886	.944	.886	.886	.944	1

Individual Brand Loyalty Scores

Our interest is to obtain s_{mi} parameters that will reveal the strength of brand loyalty for each of the consumers. Using the procedures described in the appendix, the brand loyalty scores of panel members are calculated and appear in the bottom part of Table 2. Since we have only one reference curve, the functional relation between brand loyalty and time is unidimensional; therefore, the brand loyalty scores are unidimensional. If the functional relations of the panel members were diverse such that two dimensions were necessary, we would have obtained two dimensional values for brand loyalty. Then one could classify the panel members as belonging to the first or the second reference curve or a combination of both. However, here there is only one reference curve, which means that all panel members are homogeneous regarding type of functional relation. However, their loyalty strength is heterogeneous as evident from the brand loyalty scores that range from .08623 to .31225.

The brand loyalty scores of some panel members are given below together with their purchase pattern:

Sequence	*Purchase pattern*		*Score*	*Panel member*
(1)	1 1 1 1 1 1 1	=	.31225	(1)
(2)	0 1 1 1 1 1 1	=	.26998	(5)
(3)	0 0 1 1 1 1 1	=	.22407	(3)
(4)	1 1 1 1 0 0 1	=	.22394	(10)
(5)	0 1 1 0 1 1 1	=	.22366	(14)
(6)	0 0 0 0 0 1 1	=	.08623	(2)

The brand loyalty scores immediately show the n-trial dependency. In fact, they provide values for all logical possibilities of n-trial dependencies which for seven trials number 128. In other words, the model accommodates each terminating branch of a tree diagram for a two-state n trial system. Note also that in sequences (3), (4) and (5) the purchase frequency of the favorite brand is the same, but the brand loyalty scores differ because of different arrangements. However, the degree of difference is negligible. This clearly implies that compared to frequency the influence of pattern of a given number of purchases is not large, at least for the three sequences considered.

More important, the brand loyalty scores support the a priori, theoretically based expectation of learning brand preferences. The statistical learning theory with the traditional negatively accelerating curve states that each additional consecutive purchase of a brand will add a fraction of learning still incomplete. The less the prior learning, the greater the increment. This can be seen by comparing the first differences between sequences (1), (2), and (3). The differences are .31225 − .26998 = .04227, and .26998 − .22407 = .04591. The last fraction is greater than the first because prior learning was less, as the purchase pattern shows.

Other Supporting Evidence

The brand loyalty scores were validated by two other approaches summarized here. First, tests of the null hypothesis that a two-state sequence could have been generated by chance were performed on each purchase sequence. This strongly supported the distribution of brand loyalty scores. Second, another product, namely bread, was used to provide comparison of brand loyalty scores. By the same tests, it was found that brand loyalty was lower for bread, and the brand loyalty scores supported this.[3]

[3] See [24, Chapter 3; 25] for a fuller discussion.

Table 2
REFERENCE CURVES AND PANEL MEMBERS' BRAND LOYALTY SCORES FOR RICE
($n = 7, N = 14$)

Trial	Reference curves			Runs test
	Curve I	Curve II	Curve III	
1	.71356	.68768	−.10013	Curve I:
2	.94803	.08364	.27699	$n = 7, m = 0, u = 1$ Significant
3	.98079	−.04790	.11731	Curve II:
4	.95665	.05332	−.09808	$n = 3, m = 4, u' = 4$ $p(u \leq u') = .542$
5	.94803	−.19430	.02306	Curve III:
6	.93329	−.27832	−.19046	$n = 3, m = 4, u' = 5$
7	.95750	−.13585	−.05899	$p(u \leq u') = .80$
Percentage variance explained	85.3	8.8	2.2	
Latent root	5.97	.62	.15	
Σd^2	.0563	.4804	.2918	

Panel member	Score	Panel member	Score
Brand loyalty scores			
1	.31225	8	.31225
2	.08623	9	.31225
3	.22407	10	.22395
4	.31225	11	.26998
5	.26998	12	.31225
6	.31225	13	.26998
7	.31225	14	.22366

LIMITATIONS OF THE MODEL

Despite the versatility of the factor analytic model of brand loyalty to generate more robust measures of individual buying behavior, it has several limitations.

First, the model is essentially an empirical model although in psychometrics several successful attempts have been made to use it as a theoretically based model. As an empirical model, it resembles a statistical technique more than a model since the latter is commonly used to describe a hypothesis based on some theory. This limitation, however, is not as serious as it appears. It is possible to define a priori a functional relation between time and purchase behavior based on some theory of consumer behavior, and then use the model to test whether the theoretical relation is justified in a particular situation. For example, if we assume that probability of buying in a specific situation is a stationary Bernoulli process, then the functional relation will be $y = c$ where c is the constant level of brand loyalty. Data analysis using the factor analytic model would then confirm or reject the theoretical hypothesis.

Second, brand loyalty scores obtained using the model are more difficult to interpret than probability measures. The difficulty arises because the scores do not automatically represent the probabilities of buying the brand. Although there is the unique starting point (lower limit value), namely zero value for a consumer who never buys the brand under investigation, there is no terminating point (upper limit value) defined a priori as with probability measures. The upper limit is only empirically derived from the analysis, and it varies from one investigation to another depending on the number of trials and the functional relation in each specific situation.

However, the brand loyalty scores, similar to the probability measures, tell us the relative odds of consumers' buying the same brand again. Since we are more accustomed to probability notions, an interesting extension of this research would be to establish isomorphic transformation of brand loyalty scores into probability measures. The resulting probabilities would then be functions of both frequency and pattern (history) of purchases because brand loyalty scores are themselves based on both frequency and pattern of purchases.

Finally, there is a technical limitation. Not all functional relations between time and purchase behavior can be resolved into a unique family of reference curves or dimensions. There are some nonlinear functions that cannot be exactly resolved.[4] Generally, if more than two significant dimensions are obtained in an

[4] See [24] for further discussion on this limitation.

empirical situation, it is safe to conclude that the model is inadequate for exact least-squares solution. An approximate solution to a satisfactory degree of explained variance, however, can be obtained in these situations by taking a few of the dimensions and obtaining s_{mi} values only for these dimensions.

SUMMARY AND IMPLICATIONS

The factor analytic model of brand loyalty seems useful in obtaining individual and environmental parameters for several kinds of functional relations. An added advantage is that it can work with theoretically defined functions or, more important, empirically derived functions, which is crucial when faced with analyzing the standard data collected for monitoring market conditions.

Preliminary results of a current investigation of the brand loyalty of five different products using the commercial panel data indicate that brand loyalty is stable and resembles a Bernoulli process.

An interesting extension of the model is to convert it to a multistate model in which several brands are simultaneously analyzed for several time dependencies. Analysis of the multistate extension may provide better measures of such aggregate phenomena as market share and effect of promotional activity because time-lag effects are built into the model.

An important implication of the factor analytic model of brand loyalty concerns further investigation of brand loyalty scores. The model provides a single number for each buyer that summarizes his pattern of purchase behavior over time. This is a much more desirable measure than one of average tendency. Using the brand loyalty scores we may discriminate between two or more groups having some explanatory or controlling variables. For example, if we experiment in a market in which some marketing activity is deliberately varied and if we would like to measure its effect on brand loyalty over time, the sample of people in the experimental market may be compared, based on their brand loyalty scores with a matched sample of people from the control market. Extending this, we can simultaneously compare effects of several marketing activities in different markets.

APPENDIX

Determination of Reference Curves

The data matrix Y containing n trials ($j = 1, 2, 3, \cdots, n$) and N buyers ($i = 1, 2, 3, \cdots, N$) can be resolved into factorial components or reference curves by using an important theorem in matrix algebra proved by Eckart and Young [5] and later by Young and Householder [36]. It is possible to consider the cell entries of any rectangular matrix either as projections of row variables on orthogonal axes represented by the columns or as projections of column variables on different orthogonal axes represented by the rows. The Eckart-Young theorem not only indicates how both row and column variables may be represented simultaneously by their projections on the same set of orthogonal axes but also that the set accounts for more variance than any other set of orthogonal axes.

The Eckart-Young procedure involves approximation of a matrix by another with a lower rank. It consists of analyzing a given data matrix into the product of three matrices containing a series of roots and vectors. Any complete $n \times N$ rectangular matrix ($j < i$; $j = 1, 2, 3, \cdots, n$; $i = 1, 2, 3, \cdots, N$) Y can be resolved as:

$$Y = U\Gamma W, \tag{9}$$

where

- U is an $n \times n$ orthogonal matrix of left principal vectors ($UU' = I$),
- W is an $N \times N$ orthogonal matrix of right principal vectors ($WW' = I$),
- and Γ is an $n \times N$ diagonal matrix containing γ_m roots ($m = 1, 2, \cdots r$) in the upper left section and zeroes elsewhere.

A matrix Y_r of rank $r(r < n)$ is constructed using the roots and vectors that give the best approximation of data matrix Y in the least-squares sense. The approximate matrix Y_r is constructed by taking the first r left principal vectors in U, the principal roots in Γ, and the first r right principal vectors in W. In other words,

$$Y_r = U_r\Gamma_r W_r. \tag{10}$$

If we let $U_r\Gamma_r = A$ and $W_r = S$ where A is an $n \times r$ matrix having a_{jm} elements and S is a $r \times N$ matrix having s_{mi} elements, then

$$Y_r = AS = \left(\sum_{m=1}^{r} a_{jm}s_{mi}\right). \tag{11}$$

The A matrix provides the estimates of the environmental parameters, and the S matrix provides the estimates of the person's parameters.

The matrix S consisting of r rows and N columns may be interpreted as a table giving the coordinates of the score points of N buyers in r space. The matrix $A = U_r\Gamma_r$ consisting of n rows and r columns may be interpreted as a table of r components or reference curves running over a series of n trials or time periods. Since the roots and vectors are ordered under Eckart-Young procedure ($\gamma_1 > \gamma_2 \cdots > \gamma_r$), each reference curve will explain successively less variance. The magnitude of the variance explained by each reference curve will be given by the squares of the principal roots γ_m. If all roots are taken into account, the total variance will be explained. However, it is sufficient to take the first r principal roots and vectors that will explain some satisfactory amount of variance, and the rest may be treated as resulting from random fluctuations.

Resolution of the data matrix Y into the product of three matrices is easily done if Y is multiplied by its transpose Y':

$$\begin{aligned} YY' &= (U\Gamma W)(W'\Gamma U') \\ &= U\Gamma^2 U', \end{aligned} \tag{12}$$

realizing that $WW' = I$.

The YY' matrix, which we will call the cross-products matrix, contains the sums of squares and cross products. The formation of a square symmetric matrix then transforms the problem to one of finding the characteristic roots and vectors of YY'.[5] U contains the characteristic vectors, and Γ^2 contains the characteristic roots. Note that the characteristic roots of YY' or $Y'Y$ are squares of the principal roots of Y. Each root is ordered such that the first accounts for maximum variance. Taking the first set of r characteristic roots and vectors will enable us to create the A matrix. Thus

$$A = U_r\Gamma_r .$$

The procedure just described is similar to the principal components analysis developed by Hotelling [13]. However, there are several significant differences, both in theory and in the steps involved in the two procedures. First, Hotelling's interest is to maximize the variance explained by the first component, but we minimize the error in the whole system. The two objectives do not coincide nor give the same answer. Second, Hotelling provides only an iterative process, but using the eigenvalue-eigenvector theorem in our method gives an exact solution. Third, Hotelling's method derives the components successively, each explaining the maximum variance in the residual system. The procedure described here is a simultaneous, least-squares solution. Finally, the assumptions underlying the Eckart-Young theorem are fewer and permit use of any data. The only restrictions are that the data matrix be rectangular ($n < N$) and that it be filled. This permits one to use both qualitative or quantitative or nonmetric and metric data [2]. Furthermore, it is legitimate to use a cross-products matrix as opposed to a covariance or correlation matrix. A correlation matrix is used in the standard R-type factor analysis, but it removes information about both the mean and the dispersion by standardizing to zero mean and unit variance. Such standardization is, of course, necessary if our measures on row variables are different in kind. A covariance matrix is used in principal components analysis, but it removes information about the mean. A cross-products matrix retains both the mean and the dispersion. In time-dependent situations, if there are individual differences at a given trial, it is necessary to use a cross-products matrix [23].

Estimating the Number of Significant Reference Curves

The Eckart-Young procedure of resolving a data matrix Y into the product of three matrices will usually result in several roots and vectors equal to the size of the matrix. Total variance, which is equal to the sum of squares of the diagonal values of the cross-products matrix YY', will be accounted for by the sum of the characteristic roots. Each characteristic root successively explains the residual variance of the matrix. In an empirical situation in which we have no theoretical relation postulated before the analysis, we do not know how many reference curves are needed. However, several procedures to determine the number of significant curves are possible.

First, if we plot the roots' values successively, we will usually obtain a discontinuity in the plot that can be used as a cutoff point for significant roots and corresponding vectors.

Second, the first reference curve, as mentioned earlier, approximates the mean curve. Since all the reference curves are orthogonal, the remaining reference curves will fluctuate about the zero line. Any significant trend in this fluctuation can be indicated by a runs test (see [29]). Here, the runs test may be applied to the groupings of pluses and minuses representing the fluctuations of the reference curve above and below the zero line. The systematic reference curves will determine the number of r dimensions [32].

Finally, in instances like ours in which the variables are the consecutive trials and the series of coefficients (a_{jm}) for a meaningful reference curve show the rate of change, we may obtain a quantitative statement of the curve's smoothness by taking the differences between consecutive coefficients and summing the squares of these differences, i.e., calculate Σd^2. The Σd^2 may help determine the number of reference curves since small values will be associated with the curve's smoothness.

[5] Because of the symmetry, we can premultiply Y with its transpose and get

$$\begin{aligned} Y'Y &= (W'\Gamma U')(U\Gamma W) \\ &= W'\Gamma^2 W. \end{aligned}$$

Then we can easily determine U.

REFERENCES

1. Wroe Alderson and Paul E. Green, *Planning and Problem Solving in Marketing*, Homewood, Ill.: Richard D. Irwin, Inc., 1964, 180–90.
2. Cyril Burt, "Factorial Analysis of Qualitative Data," *British Journal of Psychology (Statistical Section)*, 3 (November 1950), 166–85.
3. Jean E. Draper and Lassy H. Nolin, "A Markov Chain Analysis of Brand Preferences", *Journal of Advertising Research*, 4 (September 1964), 33–9.
4. Solomon Dukta and Lester Frankel, *Markov Chain Analysis: A New Tool of Marketers*, New York: Audits and Surveys Co., Inc., 1962.
5. Carl Eckart and Gale Young, "The Approximation of One

Matrix by Another of Lower Rank", *Psychometrika*, 1 (September 1936), 211–18.
6. William K. Estes, "The Problem of Inference from Curves Based on Group Data," *Psychological Bulletin*, 53 (March 1956), 134–40.
7. Louis A. Fourt, *Applying Markov Chain Analysis to NCP Brand-Switching Data*, Chicago: Market Research Corporation of America, 1960.
8. Ronald E. Frank, "Brand Choice as a Probability Process", *Journal of Business*, 35 (January 1962), 43–56.
9. Eugene Gilbert, *Advertising and Marketing Young People*, Pleasantville, N. Y.: Printers' Ink Books, 1957.
10. Lester Guest, "Genesis of Brand Awareness", *Journal of Applied Psychology*, 26 (December 1942), 800–8.
11. Frank Harary and Benjamin Lipstein, "The Dynamics of Brand Loyalty: A Markovian Approach", *Operations Research*, 10 (January–February 1962), 19–40.
12. Jerome D. Herniter and John F. Magee, "Customer Behavior as a Markov Process", *Operations Research*, 9 (January–February 1961), 105–22.
13. Harold Hotelling, "Analysis of Complex Statistical Variables into Principal Components", *Journal of Educational Psychology*, 24 (September 1933), 417–41.
14. John A. Howard and Jagdish N. Sheth, *A Theory of Buyer Behavior*, New York: John Wiley & Sons, (in press).
15. Alfred A. Kuehn, "An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management", Unpublished Ph.D. dissertation, Carnegie Institute of Technology, 1958.
16. ———, "Consumer Brand Choice as a Learning Process", *Journal of Advertising Research*, 2 (December 1962), 10–7.
17. Benjamin Lipstein, "The Dynamics of Brand Loyalty and Brand Switching", *Proceedings*, New York: Advertising Research Foundation, 1959, 101–8.
18. Richard B. Maffei, "Brand Preference and Market Dynamics", *Journal of Industrial Economics*, 9 (April 1961), 119–31.
19. ———, "Brand Preference and Simple Markov Processes", *Operations Research*, 8 (March–April 1960), 210–8.
20. James U. McNeal, "An Exploratory Study of the Consumer Behavior of Children", in James McNeal, ed., *Dimensions of Consumer Behavior*, New York: Appleton-Century-Crofts, Inc., 1965, 190–209.
21. Donald G. Morrison, "Stochastic Models for Time Series with Applications in Marketing", Technical Report No. 8, Program in Operations Research, Stanford University, 1965.
22. Patrick J. Robinson and David J. Luck, *Promotional Decision Making: Practice and Theory*, New York: McGraw-Hill Book Company, 1964, 216–20.
23. John Ross, "Mean Performance and the Factor Analysis of Learning Data", *Psychometrika*, 29 (March 1964), 67–73.
24. Jagdish N. Sheth, *A Behavioral and Quantitative Investigation of Brand Loyalty*, Unpublished Ph.D. dissertation, University of Pittsburgh, 1966.
25. ———, "Learning of Brand Preference at the Adult Age", *Journal of Advertising Research*, (in press).
26. ———, "A Review of Buyer Behavior", *Management Science*, 13 (August 1967), B718–56.
27. George Styan and Harry Smith, Jr., "Markov Chains Applied to Marketing", *Journal of Marketing Research*, 1 (February 1964), 50–5.
28. Murray Sidman, "A Note on Functional Relations Obtained from Group Data", *Psychological Bulletin*, 49 (May 1952) 263–9.
29. Freda S. Swed and C. Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives", *Annals of Mathematical Statistics*, 14 (March 1943), 66–87.
30. Ledyard R. Tucker, "Determination of Generalized Learning Curves by Factor Analysis", Educational Testing Service, (*mimeographed*) 1959.
31. ———, "Determination of Parameters of a Functional Relation by Factor Analysis", *Psychometrika*, 23 (March 1958), 19–23.
32. Ronald A. Weitzman, "A Factor Analytic Method for Investigating Differences between Groups of Individual Learning Curves", *Psychometrika*, 28 (March 1963), 69–80.
33. William D. Wells, "Children as Consumers", in Joseph Newman, ed., *On Knowing the Consumer*, New York: John Wiley & Sons, Inc., 1966.
34. J. W. Woodlock, "A Clue to Purchase Patterns—Markov's Mathemagic", *Sales Management*, 93 (September 18, 1964), 71–2.
35. Robert S. Woodworth and Harold Scholsberg, *Experimental Psychology*, Revised ed., New York: Holt, Rinehart and Winston, Inc., 1954.
36. Gale Young and A. S. Householder, "Matrix Approximation and Latent Roots", *American Mathematical Monthly*, 45 (March 1938), 165–71.