Title
Competitive-Component Analysis: A New Approach to Calibrating Asymmetric Market-Share Models

Journal
Journal of Marketing Research, 33(2)

ISSN
0022-2437

Authors
Cooper, Lee G
Klapper, Daniel
Inoue, Akihiro

Publication Date
1996-05-01

DOI
10.2307/3152149

LEE G. COOPER, DANIEL KLAPPER, and AKIHIRO INOUE*

Managers are unlikely to keep current with advanced developments in market-response analysis, and technical analysts often lack the marketplace knowledge of the many product categories they must track through syndicated data; this is a recipe for bad decisions. The authors present methods based on three-mode factor analysis and multivariate regression that can help both analysts and managers make better decisions regarding whether UPCs should be aggregated into brand units and, if so, how should the aggregation be done; which marketing instruments to track; and how to disentangle correlated promotional strategies. The practicality of this approach is demonstrated by an application to UPC-level data (25 UPCs, seven marketing instruments, and 156 weeks). In this example, to aggregate UPCs within a manufacturer into brand units would distort the relations between the marketing instruments and market responses. A multivariate regression from the competitive-component scores provides a methodologically sound and practical method for calibrating market response in such cases.

Competitive-Component Analysis: A New Approach to Calibrating Asymmetric Market-Share Models

The increased availability of syndicated, store-tracking data has correspondingly increased the pressure to use this valuable resource in brand planning by manufacturers and in category management by retailers. But managers who have the responsibility for planning are rarely equipped with sufficient analytical knowledge to oversee the construction of the sophisticated market-response models that are essential for those managers to do their jobs properly. Technical staffs, who are more likely to possess the analytical skills, are being asked to support more diverse brand-management teams and/or work in more product categories—to the point where their marketplace knowledge of the brand categories they analyze is stretched thin. So, the decisions involved in building sensible market-response models from scanner data require more technical knowledge than that possessed by managers and more category knowledge than that possessed by technical staffs.

Consider, for example, two decisions inevitably encountered in building market-response models from scanner tracking data. First, should universal product codes (UPCs) be aggregated into the brand units used in most analyses? And, if so, how should this aggregation be done? Key-account data for a relatively small category, such as catsup, might have four brand names plus private-label brands, but these translate into 25 different UPCs. Should we analyze the five brand units, the 25 UPCs, or something in between? If the UPCs within a brand are promoted using the same strategy and mix and if consumers respond similarly to the firm's marketing effort, then nothing will be lost by the aggregation. But how do we know this a priori? Second, in some categories some of the seven marketing instruments tracked in syndicated data (e.g., major newspaper advertisements; line ads; end-aisle, front-aisle, or in-aisle displays; in-ad coupons; price) are used much less than others. Major newspaper advertisements may be used much more frequently than line ads, but does this mean that we should

Lee G. Cooper is a professor at the Anderson School, University of California at Los Angeles. Daniel Klapper is a doctoral student at Humboldt University, Berlin, Germany. Akihiro Inoue is an assistant professor at Kwansei Gakuin University, Nishinomiya, Japan. The authors thank Nielsen Marketing Research for providing the data used in this study and Tom Belin for the S+ algorithm used to compute the General Location Model and for aiding with access to the Health Sciences Computing Facility. Much of this work was completed during Daniel Klapper's five-week visit to UCLA. Support for his visit was provided by Professor Lutz Hildebrandt, the University of Bielefeld and the Marketing Studies Center at the Anderson School. Preliminary discussions with Masao Nakanishi on the topic of integrating factor analysis and the multivariate-regression estimation procedure date back to 1986. The authors gratefully acknowledge their intellectual debt to him. They thank their UCLA colleagues, especially Donald G. Morrison, for their thoughtful comments. They also thank the editor of *JMR* and three anonymous reviewers for their careful reading and thoughtful comments.

Journal of Marketing Research
Vol. XXXIII (May 1996), 224–238

exclude line ads from the model? Only a few of the UPCs may ever employ store (in-ad) coupons. How infrequent does usage have to be, or how few UPCs have to be involved before it is tolerable to ignore a marketing instrument? These are questions marketing managers are ill equipped to resolve.

The answers to these questions have a practical impact on the task of model building. A fully extended attraction (multiplicative, competitive-interaction [MCI] or multinomial logit [MNL]) model, for example, allows for all possible cross-competitive effects between the m brands and the K marketing instruments—giving a total of ($m^2 \times K + m$) parameters to be estimated. For five brands and five marketing instruments, we must estimate 130 parameters; for 25 UPCs and seven marketing instruments, we must estimate 4400 parameters. There is a world of difference between the effort needed to estimate and interpret such models.

Marketing managers are even less well equipped to deal with the modeling issues that result from having highly correlated promotional strategies for a brand. Analysts of scanner data often see that temporary price cuts (i.e., price promotions) for leading brands are almost always announced in major newspaper advertisements. It makes sense to announce such events loudly. But under such circumstances, it may be analytically difficult to disentangle the effects of temporary price reductions and major advertisements. How to parse the components of correlated marketing strategies is a continuing puzzle.

We propose a combination of two methods that allows the data to answer the questions about aggregation of UPCs into brands, whether marketing instruments can be ignored, and how to disentangle the components of correlated marketing strategies. The first method is a special version of three-mode factor analysis (Tucker 1966), which we call *competitive-component analysis*, that portrays the competitive structure underlying the UPCs, the marketing instruments, and the time dimension—as well as the interrelations among these competitive structures. By estimating component scores for brands on *combination modes* of marketing instruments over weeks, we can gain insight into the underlying competitive dimensions and greatly reduce the number of parameters that must be estimated in the market-share model. Application of competitive-component analysis to the fairly typical case described later leads to the startling conclusion that the standard practice of aggregating UPCs into brand units distorts the underlying relation of marketing instruments to market response.

If aggregation of UPCs is problematic and/or the number of parameters to be estimated is dauntingly large, the second method we discuss can be used—a *multivariate-regression model* (Cooper and Nakanishi 1988, p. 148) from the competitive-component scores (cf. Tucker 1957). In the application described subsequently, the 25 UPCs were reduced to 12 brand components and the seven marketing instruments were reduced to five instrument components at a loss of almost no information. This approach decreased the number of parameters to be estimated from 4400 to 732 ($12 \times [12 \times 5 + 1]$) for a fully extended market-share model. Moreover, having known (inverse) transformations from the competitive components back to the original UPCs and marketing instruments makes it straightforward to compute the implied 4400 competitive parameters along with their standard errors.

We first describe the multivariate-regression approach to estimating the parameters of the fully extended attraction (cross-effects) model. Next, we present competitive-component analysis and the brand-strategy analysis it provides for understanding the coordination of the marketing mix for each competitor. We then describe how the multivariate regression can be adapted to the competitive-component scores. Finally, we present an application to the catsup market—using the General Location Model (Little and Schluchter; 1985 Olkin and Tate 1961) to impute the missing shares and prices for UPCs not sold in a particular time period.

ESTIMATING THE CROSS-EFFECTS MODEL

To develop market-response models that are useful in brand planning, we must reflect not only the differential effectiveness with which different brands execute their marketing strategies, but also the stable, cross-competitive effects. Cross-competitive effects reflect that brands differ in the degree to which they are influenced by the other brands' actions as well as the degree to which they exert influence on the other brands.

The attraction version of the cross-effects model is (cf. Cooper and Nakanishi 1988, p. 143):

$$A_{it} = \exp(\alpha_i + \varepsilon_{it}) \prod_{k=1}^{K} \prod_{j=1}^{m} f_{kt}(X_{jkt})^{\beta_{kij}}, \tag{1}$$

$$s_{it} = A_{it} / \sum_{j=1}^{m} A_{jt}, \tag{2}$$

where

A_{it} = the attraction to brand i in time period t (t = 1, 2,...,T),
α_i = the brand-specific intercept for brand i, reflecting brand loyalty and other constant components of attraction,
ε_{it} = the error,
f_{kt} = a double-subscripted function in which k controls whether an MCI or MNL form is used for a particular marketing instrument and t controls whether raw scores, exp(*z-scores*), or a *zeta-score* transformation is used on the explanatory variables in time-period t (Cooper and Nakanishi 1988, pp. 69–78),
X_{jkt} = the level of marketing instrument k of brand j in time period t, and
β_{kij} = the parameter showing the sensitivity of brand i to changes in brand j's marketing instrument k.

A straightforward approach to estimating the parameters of this market-share model is through multivariate regression (Cooper and Nakanishi 1988, p. 148).[1]

$$Y = X \cdot \mathcal{B} + E, \tag{3}$$

$$s^{*}_{it} = s_{it} / \tilde{s}_{\cdot t}, \tag{4}$$

[1]Other approaches to estimating cross-competitive effects similar to the cluster asymmetry model (Vanden Abeele, Gijsbrecht, and Vanhuele 1990) and the asymmetric, hierarchical market-share model (Foekens, Leeflang, and Wittink 1992) require an a priori structuring of the market. This may take more knowledge of the category than is available to an analyst.

where

$\tilde{s}_{\cdot t}$ = the geometric mean of s_{it} over i,
Y = the $T \times m$ matrix with elements $\{\log s^*_{it}\}$,
X = the $T \times (1 + m \times K)$ matrix $(J \mid X_1 \mid X_2 \mid \ldots \mid X_k)$,
J = the $T \times 1$ vector of ones,
X_k = the $T \times m$ matrix with elements $\{\log[f(X_{jkt})]\}$ $j = (1,2,\ldots,m)$,
$\mathcal{B}$ = the $(1 + m \times K) \times m$ matrix $(\mathcal{B}_1 \mid \mathcal{B}_2 \mid \ldots \mid \mathcal{B}_m)$,
$\mathcal{B}_i = (\alpha_i \mid \beta_{1i1} \ldots \beta_{1im} \mid \ldots\ldots\ldots \mid \beta_{Ki1} \ldots \beta_{Kim})'$, and
E = the $T \times m$ matrix of random errors $\{\varepsilon_{it}\}$.

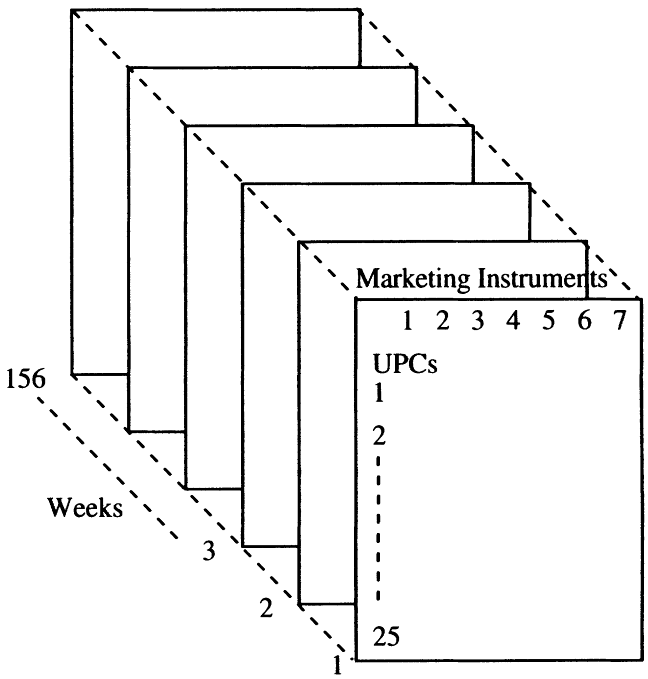
The multivariate regression greatly simplifies estimation compared with the Carpenter, Cooper, Hanssens, and Midgley (1988; hereinafter CCHM) approach. Continuing with our example of 25 UPCs, seven marketing instruments, and 156 weeks of data, the CCHM approach would construct a differential-effects matrix with 3900 rows (156 × 25) and 176 columns [1 + (7 × 25)], estimate the differential effects, cross-correlate the residuals with the marketing instruments for the other brands, add the significant cross-effects to the model, and reestimate parameters. Even if only 25% of the potential cross-effects are significant, this would add almost 1000 parameters. So the expected minimum cross-products matrix among the predictor variables would be larger than 1000 × 1000, which would need to be inverted to estimate parameters. The matrix to be inverted to solve Equation 3 is only 176 × 176, generally a much less daunting task.

We should note, however, that the 176 × 176 cross-products matrix in this example is singular, because it is based on a 156 × 176 original data matrix X. Although generalized inverses can be used in this circumstance (cf. Tucker, Cooper, and Meredith 1972), the competitive-component analysis we describe subsequently will also resolve this problem—and provide insight into the market and competitive structure as well. Using only 12 brand components and five instrument components reduces the problem to inverting a 61 × 61 cross-products matrix—a trivial task for microcomputers. Not only is the computation less onerous, but also being able to calibrate a complete cross-effects model for a singular cross-products matrix is a great asset to this approach.

COMPETITIVE-COMPONENT ANALYSIS

The basic data for brand-planning track UPCs and marketing instruments over time and form a three-way array (Figure 1). Our three-way data are described by three different modes (UPCs, Marketing Instruments, and Weeks), where the term *mode* refers to different entities that build the three-way array (e.g., Brands, Instruments, Time, Regions). The term *way* refers to the various sides (rows, columns, or slices) of the data array (Carroll and Arabie 1980).[2] This data cube holds the key to understanding market and competitive structure. If UPCs should be aggregated into brands and, if so, how UPCs should be aggregated depends on the competitive structure that underlies the brands (UPC) mode. Whether all instruments need to be included depends on the structure underlying the marketing-instruments mode. And the events that shift the nature of the competition are reflected in the structure underlying the time mode.

Figure 1
THE BASIC DATA



But a firm's strategies might include decisions to copromote certain UPCs within a brand line. Strategies might encourage coupons for large package sizes, while coordinating major ads and temporary price reductions for the more popular package sizes. Strategies for a manufacturer's own line should be known (if the analysis is being done by a manufacturer); detecting competitors' strategies, however, must come from analysis of the interactions contained within this data cube.

Numerous methods have been developed, primarily in the psychometric literature,[3] for the exploratory analysis of such three-mode, three-way data. These methods decompose the three-mode, three-way array into one, two, or three component matrices. For the analysis of the UPCs × Marketing Instruments × Weeks array, we are interested in a decomposition of all three modes simultaneously. The most general method that decomposes the three-way array into three sets of components is the Tucker3 model.[4] Tucker (1966) proposed this model for the three-mode principal components analysis that reduces the dimensionality of all three

[2]Competitive maps (Cooper 1988), for comparison, analyze a two-mode, three-way data cube (i.e., brands × brands × weeks).

[3]For surveys see the works of Kiers (1991), Kroonenberg (1992), or Law and colleagues (1984).

[4]The CANDECOMP/PARAFAC (Carroll and Chang 1970; Harshman 1970) model also analyzes three-mode three-way data but does not analyze or even allow for interactions between the UPCs and the instruments over the time. We defer discussion of this model until the end of this article.

modes to describe the information in the data. Algebraically, the Tucker3 model can be written as

$$(5)\qquad x_{ikt} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{kq} c_{tr} g_{pqr} + e_{ikt},$$

where x_{ikt} is an entry of the three-way data array with $i = 1,\ldots,m$ UPCs, $k = 1,\ldots,K$ instruments and $t = 1,\ldots,T$ time periods. In our application, a_{ip} is a coefficient that shows how strongly the i^{th} UPC is related to the p^{th} factor (or competitive component) among the UPCs. The coefficient b_{kq} shows how strongly marketing instrument k is related to the q^{th} factor (competitive component) underlying the marketing instruments. And the coefficient c_{tr} shows how strongly the t^{th} week is related to the r^{th} factor (competitive component) underlying the time mode. The elements in the core matrix G, g_{pqr}, indicate how strongly component p of brand mode interacts with component q of instruments mode and component r of time mode. The model is usually not decomposed into all possible components, but only into the first P, Q, and R components, respectively, with $P < m$, $Q < K$, and $R < T$, to provide a reduced-rank approximation. The elements e_{ikt} contain the errors resulting from the approximation.

Tucker's estimation method is based on stringing out the data matrix in three different ways and performing principal-axes analysis on each of the resulting two-way data matrices to get the component matrices A, B, and C. The components of the matrices A, B, and C are then related through the core matrix G. Because Tucker's estimation method produces only an approximate least-squares solution to the three-way array, Kroonenberg and de Leeuw (1980) developed an exact least-squares solution to the Tucker3 model (TUCKALS3) using an alternating least-squares algorithm.

After developing the general representation of the three-way array described by the modes of UPCs, instruments, and time periods, the Tucker3 model must be related to the problem of estimating differential and cross-competitive effects in a market-share model. For this purpose it is useful to express the Tucker3 model in matrix notation:

$$(6)\qquad X_t = A\left(\sum_{r=1}^{R} c_{tr} G_r\right)B' + E_t.$$

Matrix X_t is the t^{th} frontal layer of the three-way array, which contains the marketing strategies of the m UPCs for the K instruments in the t^{th} time period. G_r is the r^{th} layer of the core matrix summarizing the relation between the components of mode A and mode B for the r^{th} time component.

The Tucker3 model also can be expressed by the following representation:

$$(7)\qquad X_t = AH_tB' + E_t,$$

where

$$(8)\qquad H_t = \sum_{r=1}^{R} c_{tr} G_r.$$

H_t is the linear combination of the R frontal layers G_r of the core matrix with each component of the time period t. In his analysis of individual differences in multidimensional scaling, Tucker (1972) called H_t the *individual characteristic matrix.* H_t contains the strength of the interaction between the P UPC components and the Q instruments components for one time period. The element h_{pqt} of the matrix H_t can be interpreted as the component score of the p^{th} and the q^{th} component in the t^{th} period. Strung out into a row vector, the matrix H_t can serve as the t^{th} observation for a simplified version of the multivariate-regression model discussed previously (Tucker 1957). Now, instead of using the whole data matrix X to estimate the differential and cross-competitive effects, only the component scores for the t periods are necessary. Instead of estimating $m^2 \times K + m$ parameters, only $P^2 \times Q + P$ parameters need to be estimated. The UPCs are bundled to brand factors that represent one or more UPCs that follow a similar strategy in improving their sales and market shares. The instruments are bundled to instrument components representing an instrument strategy. From the overall fit of the model, we can judge how well the brand components and the instrument components represent the original information in the data. The fit of the Tucker3 model is derived from the relation of the explained variation to the original variation in the data:

$$(9)\qquad SS(Fit) = \frac{\sum_{i=1}^{m}\sum_{k=1}^{K}\sum_{t=1}^{T} \hat{x}_{ikt}^2}{\sum_{i=1}^{m}\sum_{k=1}^{K}\sum_{t=1}^{T} x_{ikt}^2}$$

$$(10)\qquad \hat{x}_{ikt}^2 = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{kq} c_{tr} g_{pqr}.$$

When the component matrices A, B, and C are restricted to be columnwise orthonormal, the fit of the Tucker3 model is summarized by the core matrix G, which contains all the variation in the data. Each element of the core matrix g_{pqr} indicates how much the combination of the p^{th} component of mode A, the q^{th} component of mode B, and the r^{th} component of mode C contributes to the overall fit of the model (Kroonenberg 1983, p. 158). Just as each element of the core matrix is related to the overall fit, the extended core matrix (characteristic time matrix) can be related to the fit. The sum of the squared elements of the P and Q components of $\sum_{p=1}^{P}\sum_{q=1}^{Q} h_{pqt}^2$ expresses how much variation of the t^{th} time period is explained by the characteristic matrix H_t. This expression allows for the estimation of time-specific goodness-of-fit measures.

The entries in H_t are like interaction-effect components. We also can inspect the scores of all combinations of levels of two modes on the components of the third mode—more like main-effect components. Thus, we can relate the characteristic matrix for each time period to the components of either the UPCs or the instruments. Component scores for the instruments and time periods on the brand components can be derived from the following equations:

$$(11) \qquad x_{ikt} = \sum_{p=1}^{P} a_{ip} d_{pkt} + e_{ikt}$$

$$(12) \qquad d_{pkt} = \sum_{q=1}^{Q} \sum_{r=1}^{R} b_{kq} c_{tr} g_{pqr}.$$

An element d_{pkt} can be viewed as the component score of an instrument in a time period on component p of the first mode, the UPC components. If we focus on the issue of estimating brand factors from UPCs, the elements in the matrix $A = \{a_{ip}\}$ are analogous to brand factor loadings and the elements in $D = \{d_{pkt}\}$ are analogous to brand factor scores. Similarly, the component scores of the UPCs and the time periods on the instrument components can be derived:

$$(13) \qquad x_{ikt} = \sum_{q=1}^{Q} b_{kq} f_{iqt} + e_{ikt}$$

$$(14) \qquad f_{iqt} = \sum_{p=1}^{P} \sum_{r=1}^{R} b_{kq} c_{tr} g_{pqr}.$$

If we focus on the issue of estimating instrument factors from UPCs, the elements in the matrix $B = \{b_{kq}\}$ are analogous to instrument factor loadings and the elements in $F = \{f_{iqt}\}$ are analogous to instrument factor scores. The element f_{iqt} is a component score of a UPC in a time period on the q^{th} component of mode B, the instruments mode.

Similar to the multivariate regression based on the characteristic time matrix, a regression can be based on the component scores in either matrix D (estimating $[P^2 \times K + P]$ parameters) or matrix F (estimating $[m^2 \times Q + m]$ parameters).

The next section describes the application of these methods to the analysis of store-tracking data from the catsup category.

APPLICATION TO THE CATSUP CATEGORY

The data came from the Single-Source Database provided by Nielsen Marketing Research. In one of the market areas, five stores were found that had a common promotional environment for the 25 UPCs in the catsup category. Isolating these stores provided a data set that resembled key-account data in one region. We aggregated the sales for these five stores so that the data set consists of 3006 observations over a period of 156 weeks from 1986 through 1988. Eleven variables were available:

1. week,
2. UPC,
3. dollar volume,
4. UPC unit volume,
5. line ads,
6. major ads,
7. end-aisle display,
8. front-aisle display,
9. in-aisle display,
10. store coupon, and
11. price per ounce.

Measures 5–10 are reported as the percentage of UPC volume sold when one of these promotion vehicles was in use. As already indicated, 25 different UPCs are present in this data set. They differ with respect to unit weight, package (glass, plastic, or can), special attributes (hot, tangy, lite, or no salt), and price per ounce. Twenty of the 25 UPCs belong to nationally or regionally distributed brands, and 5 UPCs are so-called private-label brands available only in that particular chain. In Table 1, we provide the UPC description and the market shares for the 25 UPCs. "Market Share" is recorded as the percentage of total volume over the 156 weeks, and price is recorded as the volume-weighted, average cents per ounce.

We see a great variation in shares, from over 20% of the market in UPC 23, a 32-ounce private-label brand, to.12% for Heinz Lite, a specialty UPC from the largest brand in the category. Four UPCs (6, 19, 21, and 23) account for over 64% of the market, and the other 21 account for less than 36%. Note that two of the largest four UPCs are private labels. These UPCs are usually aggregated regardless of their sales volume—a decision that we reevaluate subsequently. Such wide variation and the presence of eight UPCs each having less than .5% of the market also tempts many analysts to aggregate. In Cooper's (1993) study, for example, all the Heinz UPCs were combined and all the private labels were combined.

In Table 1, we also report the number of weeks during the 156-week period that each marketing-mix instrument was used for each UPC. The integer values for line ads, major ads, and store coupons reflect that these vehicles applied to all five stores in the chain. The fractional values for the various forms of in-store displays indicate that these varied from one store in the chain to another. These values aggregate the fraction of all commodity volume for each UPC that was sold using each instrument. Thus, if for three weeks one UPC sold 40% of its volume using an end-aisle display, the value would be 1.2 (3 × .4). We see that line ads are used by just four UPCs and only one week each. Is this infrequent enough to ignore? We will see if the proposed analysis can help us answer such questions.

Data Preprocessing

Data preprocessing involves the decisions of how to standardize the data and deal with missing data. The choice of how to standardize the data for three-mode analysis is an important one that does affect the solutions we obtain. Fortunately, however, the proper standardizations are dictated by the general attraction model we use here. The dependent measures are log-centered using equations 3 and 4. In this application we use exp(z-scores) to reflect the distinctiveness of each marketing action (cf. Cooper and Nakanishi 1988, pp. 69–78). To correspondingly log-center the explanatory variables in X, all we need to do is transform each variable into z-scores within each time period.

The original data set is incomplete—only 9 of the 25 UPCs have all 156 weeks of data available. The other UPCs have from 26 to 154 weeks of data available. In no single week are all 25 UPCs available. The missing promotional

Table 1

DESCRIPTIONS, MARKET SHARES, AND MARKETING MIX ACTIVITIES FOR THE 25 UPCS

Descritpion	*Weight Ounces*	*Market Share*	*Line Ad*	*Major Ad*	*End Aisle*	*Front Aisle*	*In Aisle*	*Store Coupon*	*Cents Per Ounce*
1 HEINZ KETCHUP LITE	13.25	.109	0	0	0	0	0	0	7.40
2 HEINZ KETCHUP	44.00	2.743	0	**1**	0	0	0	0	4.60
3 HEINZ KETCHUP PLS	40.00	1.815	0	**4**	0	0	0	0	5.63
4 HEINZ KETCHUP PLS	64.00	2.351	0	**3**	0	0	0	0	4.76
5 HEINZ KETCHUP	14.00	2.896	0	**1**	0	0	0	0	5.68
6 HEINZ KETCHUP	32.00	16.743	0	**19**	**3.70**	0	**2.64**	**4**	4.32
7 HEINZ KETCHUP PLS	28.00	7.754	0	**4**	**.35**	0	0	0	5.66
8 HEINZ K N ON K PLS	28.00	.496	0	**1**	0	0	0	0	6.10
9 HEINZ KETCHUP	24.00	.425	0	0	0	0	0	0	6.28
10 HEINZ KETCHUP HOT	14.00	.239	0	0	0	0	0	0	6.40
11 BROOKS CATSUP TANGY	32.00	1.749	0	**2**	0	0	**.33**	**2**	5.19
12 BROOKS CATSUP TANGY PLS	28.00	.524	0	0	0	0	0	0	5.67
13 DEL MONTE KETCHUP	32.00	3.998	0	**6**	**9.77**	0	0	0	4.28
14 DEL MONTE KETCHUP PLS	28.00	.442	0	**1**	0	0	**5.25**	0	5.06
15 HUNT'S KETCHUP NO SALT	14.00	.180	0	0	0	0	0	0	6.17
16 HUNT'S KETCHUP	14.00	.783	**1**	0	0	0	0	0	5.98
17 HUNT'S KETCHUP PLS&GLS	44.00	2.430	**1**	0	0	0	0	0	4.43
18 HUNT'S KETCHUP CAN	114.00	.219	0	0	0	0	0	0	3.21
19 HUNT'S KETCHUP PLS&GLS	32.00	14.060	**1**	**12**	**4.92**	**3.34**	**1.28**	**2**	4.16
20 HUNT'S KETCHUP PLS	17.00	.308	0	0	0	0	0	0	6.15
21 CTL BR CATSUP X FANCY	32.00	13.145	**1**	**9**	**18.34**	**1.24**	**2.65**	**1**	2.34
22 CTL BR CATSUP CAN	115.00	.215	0	0	0	0	0	0	2.43
23 CTL BR CATSUP FANCY	32.00	20.590	0	**10**	**18.16**	**3.36**	**5.99**	**4**	2.66
24 CTL BR CATSUP	14.00	1.309	0	0	0	0	0	0	4.09
25 CTL BR CATSUP FANCY PL	28.00	4.477	0	0	**3.62**	**1.45**	**.99**	**1**	4.18

instruments are easily replaced by 0s, but the missing prices and sales are another matter. First, imputing sales values of 0 is not acceptable because in general these values must be log-transformed. Second, because both sales and price data are missing, previous imputation schemes used in marketing are not appropriate. Malhotra's (1987) algorithm is developed for missing dependent measures when we at least know the sign of the missing entry, and Cooper, de Leeuw, and Sogomonian (1991) concentrate on missing independent variables.

We apply the General Location Model (Little and Schluchter 1985; Olkin and Tate 1961) to impute values for the missing prices and sales. (The procedure is described in Appendix A.) After completing the data matrix, each marketing instrument was standardized by computing z-scores within each week. This procedure transformed the variables into a form that can be directly input into the three-way analysis or directly input into the multivariate regression model (forming an MNL model in standard scores or an MCI model with exp(z-scores)).

Singular-Value Decomposition of One-Mode Two-Way Matrices

In this step of the analysis, the singular values of each mode of the data set were obtained independently to help determine the proper number of dimensions to retain. Judgments made on the successive contribution of each additional component were what Tucker (1966) used in his original procedure. The three-way data were first strung out in three different ways, and singular-value decompositions of the three different cross-product matrices were estimated.[5] The rule of thumb Tucker used for selecting dimensionality looked at the differences in the size of successive singular values. All the components prior to the last large drop in first differences were retained. This heuristic procedure gives at best an approximate answer to the question of dimensionality. The singular values of the cross-product matrix P_i (the UPC mode), where:

[5]The singular values (expressed as percentages of the trace of the corresponding cross-products matrix) and the first difference between successive singular values are listed in Appendix B.

$$(15) \qquad P_i = \{p_{ii'} = \sum_{k=1}^{K} \sum_{t=1}^{T} x_{ikt} x_{i'kt}\},$$

indicated that we should retain 12 components (containing 98.5% of the singular-value sum) to represent the UPC mode. The singular-value decomposition of the cross-product matrix P_k of the instruments, where

$$(16) \qquad P_k = \{p_{kk'} = \sum_{i=1}^{m} \sum_{t=1}^{T} x_{ikt} x_{ik't}\},$$

indicated that 5 components should be retained. The singular-value decompositions of the cross-product matrix P_t of the weeks, where

$$(17) \qquad P_t = \{p_{tt'} = \sum_{i=1}^{m} \sum_{k=1}^{K} x_{ikt} x_{ikt'}\},$$

indicated that 29 components should be retained.

In general, we expect the forces underlying marketing efforts to be at least somewhat correlated with one another. We expect price-related influences to cluster together into one or more economy factors and different kinds of newspaper features to cluster together into one or more feature factors. But we also generally expect the economy factor(s) and the feature factor(s) to be correlated, given the popular strategy of announcing low prices in newspaper features. Similarly, in the search for brand factors, we anticipate that a collection of UPCs sharing a single brand name will relate to more than one brand factor. If this occurs, we expect those brand factors to be at least somewhat correlated with each other due to copromotion, common advertising, or the brand loyalty built over time. Because of these expectations, we use an oblique rotation to represent the underlying structures. OBLIMIN, an oblique simple-structure rotation (cf. Mulaik 1972), does a good job of avoiding the problem of finding reference vectors that are too highly correlated with each other. So we use the OBLIMIN-rotated component matrices for the A (UPCs) and B (marketing instruments) modes. We are free to choose rotations for interpretive ease, because any nonsingular rotation does not affect the parameters of the multivariate regression (as is discussed subsequently in relation to Equation 27). Any nonsingular rotation can be applied to the component matrices as long as the core matrix is counter-rotated to preserve the relation to the original data. Let

$$(18) \qquad X_t = ATT^{-1}\left(\sum_{r=1}^{R} c_{tr} G_r\right)\Phi'^{-1}\Phi'B' + E_t$$

and

$$(19) \qquad X_t = A^*\left(\sum_{r=1}^{R} c_{tr} G_r^*\right)B^{*\prime} + E_t,$$

where

$A^* = AT$,
$B^{*\prime} = \Phi'B'$, and
$G^*_r = T^{-1}G_r\Phi'^{-1}$.

The values of A* are displayed in Table 2.[6] The first four components are for all practical purposes, specific to the four largest-share UPCs (see Table 1 for shares). The fifth component is the only one that looks like a common factor (i.e., multiple items loading on a single factor). The specialty UPCs (Heinz Lite and Heinz Hot) anchor the negative pole and the giant-size cans (Hunts 114 oz. can and Ctl Br 115 oz. can) anchor the positive pole. The remaining seven components are specific to seven smaller-sized UPCs, but in a manner unrelated to market share. Even though the retained components on the UPC mode accounted for 98.6% of the sum of squares, nine UPCs have no major weights ($a_{ip} > .2$) on any component. Heinz UPCs weigh heavily on five components, Del Monte UPCs on two, Hunts UPCs on two, and the five private-label UPCs on four.[7] Note that no pair of dimensions in the OBLIMIN solution is correlated as high as .05—indicating this pattern is not the result of high correlations between the oblique dimensions.

This pattern of weights casts serious doubt on the practice of aggregating UPCs with common brand names. If UPCs load on a common factor, they can be aggregated. If UPCs load on two factors, one of which is essentially constant, then the UPCs can be aggregated. But averaging (or totaling) results for brand units with a structure that spans multiple factors creates what Estes (1956) calls *functions modified by averaging.* The form of the functional relation between the explanatory variables for these aggregated brand units and the dependent measure is distorted by aggregation. The proper procedures in such cases are to calibrate the market-response model based on either the 25 UPCs or the 12 competitive-component scores. Neither approach will be distorted by UPC aggregation.

Table 3 contains the OBLIMIN rotated pattern of the marketing instruments. This pattern has no common factors. We have essentially specific components for price, end-aisle displays, major ads, in-aisle displays, and front-aisle displays. Line ads and store coupons have no major weights ($b_{kq} > .2$) on these components. So the four store-weeks of line ads are apparently few enough to ignore, but generalization of this rule is tenuous. Note that store coupons were used in 14 store weeks (see Table 1). Store coupons had no major loading, whereas front-aisle displays that had fewer store-weeks weighed heavily on Component 5. Again the pattern is not the result of highly correlated oblique components—there is no correlation higher than .20 in absolute value.

The patterns for UPCs and for marketing instruments are different from what factor analysts have come to expect in such domains as attitude or ability measurement, achievement or aptitude testing, or personality assessment, in which common factors are the rule. Perhaps we are naive to expect familiar patterns when we are analyzing such unfamiliar data. Although the excellent fit reassures us that we have retained in these components the essential variation (infor-

[6]Many of the three-mode computations were performed using the TUCKALS3 and TUCKALS2 programs by P. M. Kroonenberg (e-mail: kroonenb@rulfsw.leidenuniv.nl).

[7]This is not the result of retaining a large number of components. Preliminary research retaining five, seven, and ten components revealed conceptually similar patterns.

Table 2

FINAL COMPONENT MATRIX MODE A: OBLIMIN ROTATED AND REFLECTED

		COMP1	*COMP2*	*COMP3*	*COMP4*	*COMP5*	*COMP6*	*COMP7*	*COMP8*	*COMP9*	*COMP10*	*COMP11*	*COMP12*
1 HEINZ KETCHUP LITE	13.25	−.080	−.078	−.078	−.078	**−.415**	−.078	−.072	−.078	−.081	−.066	−.069	−.064
2 HEINZ KETCHUP	44.00	−.065	−.063	−.066	−.064	.157	−.066	−.058	−.063	−.080	−.068	−.093	−.098
3 HEINZ KETCHUP PLS	40.00	−.007	−.008	−.007	−.009	−.007	−.008	−.008	−.008	−.009	−.010	**.976**	−.008
4 HEINZ KETCHUP PLS	64.00	−.007	−.009	−.008	−.008	−.007	−.009	−.009	−.008	−.008	−.010	−.007	**.974**
5 HEINZ KETCHUP	14.00	−.075	−.074	−.076	−.075	−.059	−.075	−.074	−.073	−.094	−.107	−.097	−.124
6 HEINZ KETCHUP	32.00	−.008	−.008	**.977**	−.008	−.007	−.008	−.009	−.009	−.009	−.009	−.009	−.009
7 HEINZ KETCHUP PLS	28.00	−.008	−.010	−.009	−.009	−.010	−.010	−.010	−.010	−.010	**.961**	−.009	−.010
8 HEINZ K N ON K PLS	28.00	−.061	−.065	−.064	−.059	−.149	−.065	−.064	−.064	−.077	.035	−.064	−.065
9 HEINZ KETCHUP	24.00	−.074	−.072	−.074	−.074	−.185	−.071	−.089	−.074	−.057	−.114	−.076	−.071
10 HEINZ KETCHUP HOT	14.00	−.072	−.071	−.071	−.072	**−.213**	−.070	−.066	−.070	−.066	−.069	−.064	−.059
11 BROOKS CATSUP TANGY	32.00	−.007	−.009	−.008	−.009	−.007	−.009	−.009	−.009	**.975**	−.008	−.008	−.009
12 BROOKS CATSUP TANGY PLS	28.00	−.065	−.064	−.066	−.063	−.065	−.065	−.061	−.065	−.047	−.053	−.061	−.054
13 DEL MONTE KETCHUP	32.00	−.007	−.008	−.007	−.008	−.007	**.977**	−.009	−.009	−.008	−.008	−.007	−.008
14 DEL MONTE KETCHUP PLS	28.00	−.007	−.009	−.007	−.008	−.007	−.009	−.009	**.977**	−.009	−.008	−.007	−.008
15 HUNT'S KETCHUP NO SALT	14.00	−.078	−.077	−.079	−.085	−.164	−.077	−.089	−.079	−.114	−.161	−.083	−.092
16 HUNT'S KETCHUP	14.00	−.072	−.071	−.075	−.071	−.123	−.067	−.093	−.074	−.056	−.122	−.070	−.072
17 HUNT'S KETCHUP PLS&GLS	44.00	−.051	−.049	−.049	−.044	.183	−.052	−.033	−.049	−.029	.028	−.036	−.021
18 HUNT'S KETCHUP CAN	114.00	−.047	−.047	−.047	−.048	**.433**	−.046	−.047	−.047	−.047	−.054	−.043	−.042
19 HUNT'S KETCHUP PLS&GLS	32.00	−.008	−.009	−.008	**.977**	−.007	−.009	−.009	−.009	−.009	−.008	−.008	−.008
20 HUNT'S KETCHUP PLS	17.00	−.070	−.068	−.068	−.069	−.162	−.068	−.064	−.067	−.058	−.064	−.061	−.058
21 CTL BR CATSUP X FANCY	32.00	**.977**	−.008	−.007	−.008	−.007	−.008	−.008	−.008	−.008	−.008	−.008	−.007
22 CTL BR CATSUP CAN	115.00	−.041	−.042	−.041	−.042	**.591**	−.041	−.045	−.041	−.045	−.049	−.039	−.038
23 CTL BR CATSUP FANCY	32.00	−.007	**.977**	−.007	−.008	−.007	−.008	−.008	−.008	−.008	−.008	−.008	−.008
24 CTL BR CATSUP	14.00	−.050	−.049	−.049	−.050	**.252**	−.050	−.034	−.048	−.036	−.010	−.039	−.032
25 CTL BR CATSUP FANCY PLS	28.00	−.007	−.009	−.007	−.008	−.008	−.009	**.976**	−.009	−.009	−.008	−.007	−.008

Table 3
FINAL COMPONENT MATRIX MODE B: OBLIMIN ROTATED AND REFLECTED

	COMP1	COMP2	COMP3	COMP4	COMP5
1 Line Ads	.002	.018	−.006	−.008	.042
2 Major Ads	−.000	−.000	**.995**	−.000	−.000
3 End-Aisle Display	−.000	**1.000**	.000	−.001	−.001
4 Front-Aisle Display	.000	.000	−.000	−.000	**.997**
5 In-Aisle Display	−.000	.000	−.000	**1.000**	−.000
6 Store Coupon	−.018	−.006	.097	−.014	.066
7 Price	**1.000**	.000	−.000	−.000	.000

mation) we need to estimate market-response parameter, we must look deeper to understand competitive structure.

A Two-Mode Three-Way Analysis of UPC Strategies

To see the promotional strategy employed by each UPC, we use Tucker's (1972) generalization of the individual-differences model for multidimensional scaling (Tucker and Messick 1963). This analysis develops a common scaling space that in our case shows the amalgam relations between the ways marketing instruments are used by the various UPCs. The analysis also provides individual characteristic matrices that show how each UPC deviates from the common space. Individuals may differ in the amount of weight placed on the dimensions of the common space or by how strongly correlated common dimensions appear to be.

To obtain this solution, the original three-mode three-way data array was transformed into a two-mode three-way array with the UPC and instruments modes by the following steps:

1. The Euclidean (profile) distances between the instruments over the observed period of 156 weeks were estimated for each UPC separately.
2. The distances were converted to scalar products of vectors emanating from the centroid of the instruments by the formulae developed by Tucker and presented by Torgerson (1958, p. 258),

$$(20) \qquad \delta^*_{ikk'} = -\frac{1}{2}\left(\delta^2_{ikk'} - \delta^2_{i.k'} - \delta^2_{ik.} + \delta^2_{i..}\right).$$

Each slice of the three-dimensional matrix could be analyzed UPC by UPC, but this procedure would ignore the interactions between the UPCs. We followed a procedure outlined by Tucker (1972) for the case of multidimensional scaling of individual differences. Tucker shows that the similarities among objects expressed by scalar products for an individual (UPC) i can be decomposed in the following way:

$$(21) \qquad S_i = \Omega\Psi_i\Omega',$$

where

S_i contains the scalar products of the k = 1,...,K instruments for individual i,
Ω is the common scaling space for the marketing instruments for all individuals, and
Ψ_i is the individual characteristic matrix.

If we define a diagonal matrix W_i with the square roots of the diagonal entries in the matrix Ψ_i

$$(22) \qquad W_i^2 = \mathrm{Diag}(\Psi_i),$$

the matrix S_i can be transformed using the following formulae:

$$(23) \qquad Z_i = \Omega W_i,$$

$$(24) \qquad R_i = W_i^{-1}\,\Psi W_i^{-1},$$

and

$$(25) \qquad S_i = Z_i R_i Z_i',$$

where

R_i = the correlation matrix between the dimensions of the common space seen from the perspective of UPC i,
Z_i = the matrix of coordinate for the instruments from the point of view of UPC i, and
W_i = the matrix of weights for the common dimensions from the point of view of UPC i.

This method is sometimes called the Tucker2 model, and the TUCKALS2 algorithm (Kroonenberg and de Leeuw 1977) fits this model in a least-squares sense. To determine the dimensionality in the TUCKALS2 approach, a singular-value decomposition was used. The two-mode three-way array was strung out into a two-way matrix, and the cross-product matrix for the instruments was estimated (see Equation 16). The singular-value analysis resulted in the values reported in the first row of Table 4. We selected three dimensions for the common scaling space—accounting for 97.1% of the variance in the scalar products.[8] The common scaling space (unrotated and after a OBLIMIN rotation) is given in Table 5. The OBLIMIN is simpler to interpret than the unrotated, common scaling space. The first dimension strongly emphasizes price—with a slight contrast of price with front-aisle and in-aisle displays. The second dimension strongly emphasizes end-aisle displays—with a slight contrast of line ads and store coupons. The third dimension emphasizes major ads—again with a slight contrast of both front-aisle and in-aisle displays. These are the common building blocks from which the marketing-mix strategies for each UPC are constructed.

Singular-value analyses were undertaken for each individual UPC. The results are shown in Table 4, with the bold type indicating the number of dimensions we judged appropriate in each case.

Ten of the 25 UPCs are one-dimensional in their marketing strategies. In each of these cases, the single factor involved price alone. Five of these ten were the UPCs we already indicated weighed most heavily on Component 5 in the UPC-mode analysis summarized in Table 5 (i.e., UPCs 1, 10, 18, 22, and 24). Four of the remaining five price-strategy UPCs have weights between −.15 and −.19 on

[8]In the Tucker2 model, two of the component modes are supposed to produce identical matrices. When four dimensions were retained (the number that seems subjectively correct to us) and this method applied, the two component matrices were only identical for the first two components. The third and the fourth components were different. The meaning of such an outcome has not been discussed in the literature. We believed it would be prudent to select the three-dimensional solution that had converged properly.

Table 4
THE SINGULAR VALUES (SV) FOR THE 25 SCALAR-PRODUCT MATRICES

	SV1	*SV2*	*SV3*	*SV4*	*SV5*	*SV6*	*SV7*
Common	72.86	12.57	**6.47**	5.83	1.86	.41	.00
UPC1	**98.51**	.61	.31	.27	.17	.08	.02
UPC2	55.99	**31.54**	5.96	3.15	1.99	1.03	.31
UPC3	55.78	**39.38**	2.45	1.16	.72	.36	.11
UPC4	77.16	**13.89**	4.00	2.42	1.48	.78	.23
UPC5	64.26	**28.95**	3.25	1.71	1.08	.54	.17
UPC6	52.49	18.55	15.23	**10.62**	2.95	.11	.01
UPC7	52.34	**41.74**	3.52	1.18	.72	.35	.11
UPC8	**90.61**	5.04	2.17	1.05	.66	.33	.10
UPC9	**95.62**	1.73	.94	.83	.51	.26	.08
UPC10	**96.05**	1.60	.84	.73	.45	.23	.07
UPC11	44.85	31.44	**14.53**	5.97	2.48	.59	.09
UPC12	**87.37**	5.45	2.59	2.22	1.40	.71	.22
UPC13	60.69	27.90	**10.34**	.59	.30	.13	.04
UPC14	84.85	8.83	**4.33**	1.02	.58	.28	.08
UPC15	**94.64**	2.43	1.03	.92	.57	.29	.09
UPC16	77.89	**16.85**	1.95	1.24	1.04	.67	.33
UPC17	49.61	**37.50**	5.52	2.74	2.43	1.39	.78
UPC18	**96.94**	1.26	.65	.55	.34	.17	.05
UPC19	38.29	25.33	**15.31**	7.85	6.37	4.59	2.23
UPC20	**94.47**	2.22	1.19	1.02	.64	.32	.10
UPC21	57.69	18.35	**10.19**	6.23	5.03	1.30	1.19
UPC22	**98.50**	.62	.31	.27	.17	.08	.02
UPC23	49.25	15.96	12.76	10.63	**7.95**	3.42	.00
UPC24	**89.45**	3.98	2.38	2.04	1.27	.65	.20
UPC25	42.03	20.92	16.32	12.98	**7.08**	.59	.05

Component 5 (i.e., UPCs 8, 9, 15, and 20) with no larger weight (in absolute value) on any other component. It seems clear now that Component 5 represents UPCs that employ a price-dominated strategy. Referring back to the original description of marketing-mix activity in Table 1, we note that nine of the ten UPCs we identified are the only UPCs that employed no marketing instrument other than price. The tenth UPC (UPC 8) had one week of a major ad as the only marketing-mix activity other than price variation.

Seven of the UPCs in Table 4 employed two-dimensional marketing strategies. Five of the seven (UPCs 2, 3, 4, 5, and 7) are Heinz brands that use a strategy that emphasizes price and major ads. The other two are Hunts brands (UPCs 16 and 17) that weight price heavily and end-aisle displays slightly.[9]

Five UPCs have three-dimensional strategies: UPCs 13, 19, and 21 have strong weights on price, major ads, and end-aisle displays and use these marketing instruments actively. UPC 11 weights price and major ads much more heavily than end-aisle display. UPC 14 has modest weights that emphasize major ads over price and price over end-aisle displays.

Heinz 32-ounce (UPC 6) has a 16.7% share by itself. Its strategy is four-dimensional—employing all marketing instruments actively, except for line ads and front-aisle displays. Private-label UPCs 23 and 25 have five-dimensional strategies. UPC 23 is the single largest seller, holding 20.6% of the market. Every marketing instrument except line ads is used. UPC 25 has a 5.5% share. This UPC has proportionally fewer promotions but still employs all marketing instruments, except line and major ads.

The Tucker2 model has provided insight into the dimensionality and structure of the marketing strategies down to the UPC level. We find that the raw marketing-mix activities (Table 1) are a less-than-perfect guide to understanding the marketing strategies even within a brand line. We also have gained insight into the structure underlying the UPC mode displayed in Table 2. We now turn to estimating the relations between the marketing-mix activity (on which all prior analyses are based) and the market response.

A Multivariate Regression from Competitive Components

We are working with the results of a three-mode analysis that provided a solution with $P = 12$ components in mode A (UPCs), $Q = 5$ components in mode B (instruments), and $R = 29$ components in mode C (time). The $12 \times 5 \times 29$ component solution yields a fit of 94.2%, which indicates that the components account for almost all the variation in the data and that the solution is a good basis for further analysis. Tucker (1972) emphasizes looking for the last drop in singular values (or eigenvalues). If we look just for large drops (rather than the last large drop) and accept a lower-dimensional solution ($P = 2$, $Q = 2$, and $R = 5$) the fit is only 60.1%.

As outlined previously, the component scores of the week or time characteristic matrices were estimated by Equation 8. The matrix H_t was the t^{th} case in the multivariate-regression model and the t^{th} row in the regression matrix H, which has $P \times Q + 1$ columns, based on the P brand components, the Q instrument components, and a column of ones for the

[9]Remember that we analyzed z-scores. So even though these UPCs are not shown in Table 1 as using end-aisle displays, there could be a strategy of changing price when other Hunts UPCs (such as UPC 19) use end-aisle displays. Such coordination would be expected to show up as something other than a pure price strategy.

Table 5
THE TUCKALS2 SOLUTION FOR THE SCALAR-PRODUCT MATRICES

Common Scaling Space—Unrotated.			
Line Ads	−.039	−.140	**.288**
Major Ads	.078	**−.555**	**−.727**
End-Aisle Display	**.579**	**.659**	**−.263**
Front-Aisle Display	.065	−.089	**.359**
In-Aisle Display	.135	−.072	**.316**
Store Coupon	−.023	**−.222**	**.227**
Price	**−.796**	**.419**	**−.200**
Common Scaling Space—Oblimin Rotated and Reflected.			
Line Ads	−.157	**−.256**	−.204
Major Ads	−.031	−.027	**.910**
End-Aisle Display	−.027	**.910**	−.025
Front-Aisle Display	**−.244**	−.196	**−.287**
In-Aisle Display	**−.272**	−.126	**−.253**
Store Coupon	−.185	**−.279**	−.112
Price	**.915**	.026	−.028

brand-specific effects. Because the independent variables have been reduced to component scores, the dependent measure had to be related to the component matrix A. In the original data set, the dependent measures are in a matrix Y of order T × m, with i = 1,...,m UPCs (see equations 3 and 4). The matrix Y can be related to the TUCKALS3 solution by the following transformation:

(26) $$ {}_{T}Y_{P}^{*} = {}_{T}Y_{m} \cdot {}_{m}A_{P}. $$

With the matrices Y* and H, we obtain the following multivariate-regression model:

(27) $$ Y^{*} = H\beta_{PQ} + E, $$

where β_{PQ} contains the regression parameters for the P brand components, the Q instrument components, and the brand-specific intercepts. The subscripts PQ indicate the number of parameters estimated for the brand components (P) and the instrument components (Q). Equation 27 also should make it clear that the choice of component rotation is purely for interpretive ease, as H (and consequently β_{PQ}) is unaffected by any rotation of the UPC, instrument, or time-period modes.

As outlined by Tucker (1966, p. 289), the rank of the matrix H will be the rank of the component matrix C plus one for the intercept. The component matrix C is a full column-rank matrix with rank 29. To estimate the coefficient matrix β_{PQ}, we performed a full-column rank singular-value decomposition on the regression matrix H:

(28) $$ H = U\Delta V'. $$

As expected, only 30 singular values in the matrix D were different from zero. The regression coefficient matrix β_{PQ}^{*} can be obtained from the first 30 eigenvectors of U:

(29) $$ \beta_{PQ}^{*} = (U_{30}'U_{30})^{-1}U_{30}'Y^{*} = U_{30}'Y^{*}. $$

An unbiased estimate of the coefficient matrix β can be obtained from

(30) $$ \beta_{PQ} = V_{30}\Delta_{30}^{-1}\beta_{PQ}^{*}. $$

The variance-covariance matrix of the residuals and the variance of the β coefficients are:

(31) $$ \begin{aligned} \Sigma_{PQ} &= \frac{1}{n}(Y^{*} - H\beta_{PQ})' \cdot (Y^{*} - H\beta_{PQ}) \\ &= \frac{1}{n}(Y^{*} - U_{30}\Delta_{30}V_{30}'V_{30}\Delta_{30}^{-1}\beta_{PQ}^{*})' \\ &\quad \cdot (Y^{*} - U_{30}\Delta_{30}V_{30}'V_{30}\Delta_{30}^{-1}\beta_{PQ}^{*}) \\ &= \frac{1}{n}(Y^{*} - U_{30}\beta_{PQ}^{*})' \cdot (Y^{*} - U_{30}\beta_{PQ}^{*}), \end{aligned} $$

and

(32) $$ \mathrm{Var}(\beta_{PQ}) = \mathrm{Diag}(H'H)^{-1} \otimes \Sigma_{PQ}, $$

where ⊗ is the Kronecker product. The subscripts on the matrices Σ_{PQ} and β_{PQ} indicate that the matrix of errors and the matrix of regression coefficients pertain to a solution with P brand components on Q instrument components.

Although β_{PQ} are the regression estimates for the component scores of matrices A and B of the TUCKALS3 solution and could be interpreted directly, we also are interested in the regression estimates for the original UPCs on the Q instrument components. The estimation of the parameters β for the 25 UPCs from the five instrument components and the standard errors can be derived from the following procedure (the intercept-term is dropped from consideration): First the matrix β_{PQ} must be related back to the original m UPCs. The information for the inverse transformation is provided by component matrix A. Let I indicate an identity matrix and A^{-} indicate the generalized inverse of A:

(33) $$ \beta_{mQ} = (I_{Q} \otimes A) \cdot (\beta_{RPQ} \cdot A^{-}), $$

where the indices mQ on the regression matrix indicate that this matrix contains the coefficients for the m UPCs on the Q instrument components, and the β_{RPQ} are the regression coefficients for P UPC components and the Q instrument components with the R indicating that the brand-specific coefficients removed from β.[10]

The regression matrix H column for brand-specific intercepts removed are denoted as H_R and can be related to the original m UPCs using

(34) $$ X_{Q} = H_{R} \cdot (I_{Q} \otimes A^{-}). $$

The corresponding variance-covariance matrices can be estimated as

(35) $$ \Sigma_{mQ} = \frac{1}{n}(Y - X_{Q}\beta_{mQ})' \cdot (Y - X_{Q}\beta_{mQ}) $$

and

(36) $$ \mathrm{Var}(\beta_{mQ}) = \mathrm{Diag}(X_{Q}'X_{Q})^{-1} \otimes \Sigma_{mQ}. $$

[10]Equation 33 does not provide an unbiased estimate of β_{mQ}. We could get an unbiased estimate if instead of using H, we use the extended core in Equation 27.

The subscript Q indicates that the regression matrix refers to the Q instrument components.

Furthermore, we can transform the regression parameters on the Q instrument components back to the original K marketing instruments. The information for this transformation is provided by the components matrix B of the instruments:[11]

$$\beta_{mK} = (B \otimes I_m) \cdot \beta_{mQ}. \tag{37}$$

The corresponding regression matrix X can be obtained as follows:

$$X_K = X_Q \cdot (B^- \otimes I_m), \tag{38}$$

where B^- is the generalized inverse of B. The standard errors of the fully extended cross-effects model are given by

$$\Sigma_{mK} = \frac{1}{n}(Y - X_K\beta_{mK})' \cdot (Y - X_K\beta_{mK}) \tag{39}$$

and

$$\mathrm{Var}(\beta_{mK}) = \mathrm{Diag}(X'_K X_K)^{-1} \otimes \Sigma_{mK}. \tag{40}$$

The subscript K on the regression matrix X indicates that this matrix belongs to the original K instruments.[12]

Because we imputed the conditional means for missing values, we must adjust the variance-covariances of residuals. According to Little and Rubin (1987, pp. 25, 44), the following adjustment should be performed:

$$\hat{\sigma}_{ij} = \hat{\sigma}^*_{ij} \left(\frac{n - p}{m - p}\right) \tag{41}$$

where $\hat{\sigma}^*$ is the sample covariance estimated on the basis of n total samples, and m is the total number of observed samples. This implies we can either multiply the variance-covariance matrix of parameters by 1.393 = (3900 – 732)/(3006 – 732) or divide the t-value matrix of parameters by $1.180 = \sqrt{1.393}$ and infer the statistical significance as usual. Alternatively, we can simply adjust the critical values we use for testing statistical significance—multiplying the standard critical value by 1.18. In this case, we would use 1.94 instead of 1.645 (5% one-tailed α with infinite degrees of freedom), or 2.31 instead of 1.96 (5% two-tailed α with infinite degrees of freedom).

We use the adjusted one-tailed critical value (1.94) for differential (self) effects, because we typically have natural hypotheses regarding the direction of these effects (i.e., price has a negative differential [self] effect and all other instruments have positive differential [self] effects). But we use the nondirectional value for cross-competitive effects because they represent the deviation of a cross-effect from what we expect under the symmetric market hypothesis. The symmetric market hypothesis basically asserts that when one brand looses attraction, the other brands gain strictly in proportion to their market shares (cf. Bell, Keeney, and Little 1975). In analyzing a number of UPCs within a single brand line, we do not expect the symmetric market hypothesis to hold. What our cross-effects show is whether a particular effect is either more or less competitive than is implied by the symmetric market hypothesis and thus has no expected direction.

Overall, 31 of the 175 differential effects were significant (18% significant at α = .05). Of the remaining 144 differential effects, four were extreme enough to be significant in the wrong direction (2.8% wrong-signed and significant at α = .05). Three of the bad signs were for the seldom-used line ads, and none of these involved a UPC that actually used line ads. The other bad sign involved in-aisle display, which again the UPC did not use. All results involving differential effects are far better than we could expect by random chance alone. Among the major marketing instruments (price, major ads, and end-aisle displays), 15 of 75 differential effects were significant (20% significant at α = .05), and there were no bad signs.[13]

Approximately 14% of the 4200 potential cross-effects were significant at α = .05. This is certainly many more than we expect by random chance alone. Over all marketing instruments, these effects were split relatively evenly between those indicating greater competition than the symmetric market hypothesis would imply and those indicating greater cooperation. If we limit our count to the cross-effects for the three major instruments and further limit it to just the between-brand cross-effects, then 60% are competitive cross-effects and 40% are cooperative cross-effects. Still, we expect competitive cross-effects even within a brand line if not all the UPCs are copromoted simultaneously. If a display does its job and not all UPCs within a brand line are on the display, we expect cannibalization as well as competitive effects on rivals.

The overall fit of the multivariate-regression models can be assessed using the matrix correlation, RV (Ramsay, ten Berge, and Styan 1984), between the predicted and original values in the Y matrix. We can assess this correlation for each of the three multivariate regressions performed. For the 12 brand components and five instrument components, the correlation is .993; for the 25 UPCs and the five instrument components, it is .985; and for the 25 UPCs and the original seven marketing instruments, it is .985. As has been found in the other applications, this style of market-share model fits extremely well.

For comparison we estimated an approximation to the fully extended model on the original scores[14] and a model based on the CCHM specification, of asymmetric attrac-

[11] Equation 37 does not provide an unbiased estimator.

[12] An anonymous referee pointed out that these standard errors would greatly overstate the certainty that can be ascribed to OLS coefficients, because they are valid only if the market response is uncorrelated with the collinear influences in the excluded components. To test whether this condition applied to the current illustration, a multivariate regression was run on all the excluded components (UPC components 13–25, instrument components 6 and 7, and time components 30–156). The RV for this model was .054. This indicates that the excluded components were relatively unrelated to the market response. In the current case, the standard errors do not seem to overstate our certainty about the regression parameters. We thank the referee for this thought-provoking comment.

[13] The complete tables are available on request.

[14] Because the fully extended model is singular, we first eliminated all cross-effects for marketing instruments that were unused by the corresponding UPC and then approximated the needed inverse using a generalized inverse based on the nonzero singular values. This provides a solution that is usable for comparison purposes at least.

tions. The RV for the approximate fully extended model is .992—just lower than the solution based on competitive component scores. The CCHM specification fit is .989. So, the fit of the model based on competitive components is much closer to the calibration fit of the fully extended model than to that of the CCHM specification.

A validation study was conducted splitting the data into 104 weeks for calibration and 52 weeks for cross-validation. When the same decisions on number of factors were employed on this smaller data set, the cross-validity RV was .977. The study also showed that the structures underlying these data are generally robust. For a corresponding cross-validation on the approximate fully extended model, the cross-validity RV was .933 and for the CCHM specification, the cross-validity fit was .924. So the competitive component model holds up better in cross-validation than does the fully extended model or the CCHM specification.

CONCLUSIONS

We present a combination of methods that enable us to calibrate and cross-validate fully extended cross-effects market-share models from UPC-level data. We included all the marketing instruments tracked by syndicated data, even though some of the instruments were rarely used in the marketing mix of any UPC. Given almost 700 significant cross-effects, the standard approach would have required inverting an approximately 750 × 750 cross-products matrix—a major undertaking. The current approach inverts a 61 × 61 matrix —a trivial task on modern PCs. All the computations for the three-mode analysis and the multivariate regression were computed using a 20 megahertz 386 PC—old technology by now. Demonstrating the practicality of this approach makes it easier for technical staffs to undertake such analyses in categories with which they are not familiar. What guides decision making is the analysis, rather than an analyst's marketplace knowledge regarding the brands in the category.

We also developed methods for investigating the structure underlying marketing-mix strategies for each UPC. All the unidimensional strategies were dominated by price. Most of the two-dimensional strategies involved price and major ads. We saw higher-dimensional strategies for the largest UPCs. Even within a single brand line, the different UPCs were found to vary in strategy. Marketing managers should think through how multiple strategies can create synergies that enhance cooperation and minimize cannibalization.

Our effort is limited in at least two important ways. First, we cannot generalize the applicability of our new approach on the basis of the catsup category alone. For one thing, the catsup category has relatively few UPCs. The only other major work at this level of disaggregation is Fader, Hardie, and Walsh's (1994) study involving 58 UPCs. We must conduct this style analysis in more categories to see if the unusual factor patterns are typical for this style data. Second, we did not perform multiple imputations of the missing data. Although this would require corresponding multiples of three-mode components analysis and multiples of the multivariate regression, the effort would enable the investigation of the statistical properties of the parameter estimates. This is a prime candidate for further research.

We could question if the particular combination of methods we use here is the best. After all, we do use a lot of methods: the Tucker3 model, using OBLIMIN rotation and the General Location Model for missing data, followed by a multivariate regression from the competitive-component scores. Any claim to being the *best* is based first and foremost on the generality of the Tucker3 model. Three-mode models such as the CANDECOMP/PARAFAC (Carroll and Chang 1970; Harshman 1970) or CANDELINC (Carroll, Pruzansky, and Kruskal 1980) require the same number of factors to underlie UPCs, marketing instruments, and time. We have seen that different numbers of factors are needed for these different modes—12 for UPCs, 5 for marketing instruments, and 29 for time periods. These models also allow no interactions between the modes (i.e., the frontal slices of the core matrix must be diagonal). Such restrictions seem substantively inappropriate in our general case. Moreover, Kiers (1991) has demonstrated that because of these restrictions, none of the constrained models can fit better than the corresponding Tucker3 model.

Once we accept the Tucker3 model as most appropriate for representing general three-mode three-way data, the other choices are not as controversial. The General Location Model simply provides the maximum-likelihood estimate of the appropriate conditional mean for each missing sales or price. No other method has been advanced in the marketing literature that is equally advantageous for both dependent and independent measures. The choice of OBLIMIN is merely the choice of oblique simple structure as opposed to an orthogonal approximation to simple structure as in VARIMAX. We have no reason to expect factors underlying UPCs to be orthogonal. Throughout this exploration we have emphasized methods that would enable us to see what is there, as opposed to constraining the structure to fit one preconceived notion or another. Restrictions in the other models of the sorts we describe here imply a particular direction for an analysis. Consequently they violate one of the basic principles of leadership: If you do not know where you are going, do not lead.

Methodologically, we have always known where we were going. The goal was to have a series of methods that would enable us to explore the causal data underlying market response and connect the results of that exploration to the parameters of a logically consistent market-share model. To do this right, we had to maintain the metric quality of the original data. Throughout all the transformations involved in the methods we illustrate here, we have done so. The multivariate regression achieves this final goal.

In summary, we have presented methods that make analysts more able to do their job and provide marketing managers a more fine-grained picture of strategies and competition within a brand line and between brand lines. Our findings indicate it is a mistake to aggregate UPCs into brand units. Such aggregation will distort the relation of the brand's actions to the market response. A multivariate regression from competitive components, as discussed here, provides a viable alternative that avoids this aggregation problem altogether.

Appendix A

APPLICATION OF THE GENERAL LOCATION MODEL

To simplify our application of the General Location Model (Little and Schluchter 1985; Olkin and Tate 1961) we represented all the marketing-mix instruments, except price, as categorical (binary) variables. For the two continuous variables (price and sales), we assume a bivariate-normal distribution, with a mean that varies with each different combination of the categorical variables. With six categorical variables for each of 25 UPCs, we have, in essence, 1600 ($= 2^6 \times 25$) cells, each of which has bivariate normal-density for the two continuous variables (price and sales). The General Location Model posits a multinomial mixture of multivariate-normal distributions in which the within-cell covariance matrix is assumed to be the same across all the cells. For categorical missing values, the logit model is applied to impute values, and for continuous missing values, the conditional normal regression model is used.

It should be noted that, whenever we had missing prices and sales, the other six marketing instruments should have 0 values (i.e., if there were no sales for a UPC in a week, we feel justified in assuming that there were no major ads, line ads, displays, or store coupons). Hence, we imputed 0 for missing marketing instruments. Thus, we had only to impute values for the missing prices and sales. With the General Location Model, we have the ability to create multiple imputations (i.e., multiple data sets that vary in the values imputed for the missing elements). Such multiple imputations are valuable in developing bootstrap estimates for regression applications but are burdensome when an analysis such as we are undertaking follows the imputation. So we chose to complete our data matrix only once. This makes our application of the General Location Model equivalent to imputing the maximum-likelihood estimates of conditional means for missing prices and sales (e.g., Little and Rubin 1987).

Appendix B

SELECTING DIMESIONALITY FOR 3-MODE ANALYSIS

	UPCs		*Instruments*		*Time*	
	Percentage of Trace	*1st Difference*	*Percentage of Trace*	*1st Difference*	*Percentage of Trace*	*1st Difference*
1	45.07	29.91	44.22	23.72	44.85	37.70
2	15.16	5.01	20.50	6.46	7.15	.76
3	10.15	.80	14.04	4.05	6.39	1.69
4	9.35	1.97	9.99	3.10	4.71	.78
5	7.38	3.91	6.89	3.53	3.92	.83
6	3.47	.79	3.36	2.38	3.10	.18
7	2.68	1.04	.98		2.92	.20
8	1.64	.20			2.71	.48
9	1.44	.57			2.23	.11
10	.87	.11			2.12	.19
11	.76	.25			1.93	.19
12	.51	.18			1.74	.22
13	.33	.04			1.52	.26
14	.29	.03			1.26	.05
15	.26	.03			1.21	.06
16	.23	.13			1.15	.15
17	.10	.03			1.00	.09
18	.07	.00			.91	.03
19	.07	.03			.88	.11
20	.04	.03			.77	.06
21	.01	.00			.71	.06
22	.01	.01			.65	.07
23	.00				.57	.02
24					.55	.02
25					.53	.01
26					.52	.11
27					.41	.02
28					.38	.06
29					.32	.06
30					.26	.01
31					.25	.03
32					.23	.01
33					.22	.01
34					.21	.02
35					.19	.01
36					.18	.01
37					.17	.01
38					.16	.03
39					.13	.01
40					.12	.01
41					.11	.02
42					.09	.02
43					.07	.00
44					.07	.01
45					.06	.01
46					.05	.01
47					.05	.01
48					.04	.00
49					.04	.01

REFERENCES

Bell, David E., Ralph L. Keeney, and John D. C. Little (1975), "A Market Share Theorem," *Journal of Marketing Research*, 7 (May), 136–41.

Carpenter, Gregory A., Lee G. Cooper, Dominique M. Hanssens, and David F. Midgley (1988), "Modeling Asymmetric Competition," *Marketing Science*, 7 (4), 393–412.

Carroll, J. Douglas and Phipps Arabie (1980), "Multidimensional Scaling," in *Annual Review of Psychology*, Vol. 31, M. R. Rosensweig and L. W. Porter, eds. Palo Alto, CA: Annual Reviews, 607–49.

——— and Jih-Jie Chang (1970), "Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, 35 (September), 283–320.

———, Sandra Pruzansky, and Joseph B. Kruskal (1980), "CANDELINC: A General Approach to Multidimensional Analysis of Many-Way Arrays with Linear Constraints on Parameters," *Psychometrika*, 45 (1), 3–24.

Cooper, Lee G. (1988), "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (6), 299–319.

——— (1993), "Market-Share Models," in *Handbooks in Operations Research and Management Science: Marketing*, J. Eliashberg and G. L. Lilien, eds. Amsterdam: North-Holland, 259–314.

———, Jan deLeeuw, and Arum G. Sogomonian (1991), "An Imputation Method for Dealing With Missing Data in Regression," *Applied Stochastic Models and Data Analysis*, 7 (3), 213–35.

——— and Masao Nakanishi (1988), *Market-Share Analysis*. Boston: Kluwer Academic Publishers.

Estes, William K. (1956), "The Problem of Inference from Curves Based on Group Data," *Psychological Bulletin*, 53 (2), 134–40.

Fader, Peter S., Bruce G. S. Hardie, and John W. Walsh (1994), "Modeling Consumer Choice Among SKUs," Wharton Marketing Department Working Paper #94-007, February.

Foekens, E. W., P. S. H. Leeflang, and D. R. Wittink (1992), "Asymmetric Market Share Modeling with Many Competitive Items Using Market Level Scanner Data," working paper, University of Groningen.

Harshman, R. A. (1970), "Foundations of the PARAFAC Procedure: Models and Conditions for an 'Explanatory' Multi-Mode Factor Analysis," *UCLA Working Papers in Phonetics*, 16, 1–84.

Kiers, H. A. L. (1991), "Hierarchical Relations Among Three-Way Methods," *Psychometrika*, 56 (September), 449–70.

Kroonenberg, P. M. (1983), *Three-Mode Principal Component Analysis*. Leiden, The Netherlands: DSWO Press.

——— (1992), "Three-Mode Component Models: A Survey of the Literature," *Statistica Applicata*, 4, 619–33.

——— and J. de Leeuw (1977), "TUCKALS2: A Principal Component Analysis of Three-Mode Data," Research Bulletin RB. 001-77, Department of Data Theory, University of Leiden, Leiden, The Netherlands.

——— and ——— (1980), "Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms," *Psychometrika*, 45 (March), 69–97.

Law, H. G., C. W. Snyder Jr., J. A. Hattie, and R. P. McDonald (1984), *Research Methods for Multimode Data Analysis*. New York: Praeger Press.

Little, R. J. A. and D. B. Rubin (1987), *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

——— and M. D. Schluchter (1985), "Maximum Likelihood Estimation for Mixed Continuous and Categorical Data with Missing Values," *Biometrika*, 72 (3), 497–512.

Malhotra, Naresh K. (1987), "Analyzing Marketing Research Data With Incomplete Information on the Dependent Variable," *Journal of Marketing Research*, 24 (February), 74–84.

Mulaik, S. (1972), *The Foundations of Factor Analysis*. New York: McGraw-Hill.

Olkin, I. and R. F. Tate (1961), "Multivariate Correlation Models with Mixed Discrete and Continuous Variables," *Annals of Mathematical Statistics*, 32 (June), 448–65.

Ramsay, J. O., J. ten Berge, and G. P. H. Styan (1984), "Matrix Correlation," *Psychometrika*, 49, 403–23.

Torgerson, W. S. (1958), *Theory and Methods of Scaling*. New York: John Wiley & Sons.

Tucker, Ledyard R (1957), "Transformation of Predictor Variables to a Simplified Regression Structure," *Technical Report*. Princeton, NJ: Educational Testing Service.

——— (1966), "Some Mathematical Notes on Three-Mode Factor Analysis," *Psychometrika*, 31 (September), 279–311.

——— (1972), "Relations Between Multidimensional Scaling and Three-Mode Factor Analysis," *Psychometrika*, 37, 3–27.

———, Lee G. Cooper, and William Meredith (1972), "Obtaining Squared Multiple Correlations from a Correlation Matrix Which May Be Singular," *Psychometrika*, 37 (2), 143–48.

——— and Samuel Messick (1963), "An Individual Differences Model for Multidimensional Scaling," *Psychometrika*, 28 (December), 333–67.

Vanden Abeele, P., E. Gijsbrecht, and M. Vanhuele (1990), "Specification and Empirical Evaluation of a Cluster-Asymmetry Market Share Model," *International Journal of Research in Marketing*, 7 (4), 223–47.