24

MODELING MARKETING MIX

GERARD J. TELLIS
University of Southern California

CONCEPT OF THE MARKETING MIX

The marketing mix refers to variables that a marketing manager can control to influence a brand's sales or market share. Traditionally, these variables are summarized as the four Ps of marketing: product, price, promotion, and place (i.e., distribution; McCarthy, 1996). Product refers to aspects such as the firm's portfolio of products, the newness of those products, their differentiation from competitors, or their superiority to rivals' products in terms of quality. Promotion refers to advertising, detailing, or informative sales promotions such as features and displays. Price refers to the product's list price or any incentive sales promotion such as quantity discounts, temporary price cuts, or deals. Place refers to delivery of the product measured by variables such as distribution, availability, and shelf space.

The perennial question that managers face is, what level or combination of these variables maximizes sales, market share, or profit? The answer to this question, in turn, depends on the following question: How do sales or market share respond to past levels of or expenditures on these variables?

PHILOSOPHY OF MODELING

Over the past 45 years, researchers have focused intently on trying to find answers to this question (e.g., see Tellis, 1988b). To do so, they have developed a variety of econometric models of market response to the marketing mix. Most of these models have focused on market response to advertising and pricing (Sethuraman & Tellis, 1991). The reason may be that expenditures on these variables seem the most discretionary, so marketing managers are most concerned about how they manage these variables. This chapter reviews this body of literature. It focuses on modeling response to these variables, though most of the principles apply as well to other variables in the marketing mix. It relies on elementary models that Chapters 12 and 13 introduce. To tackle complex problems, this chapter refers to advanced models, which Chapters 14, 19, and 20 introduce.

The basic philosophy underlying the approach of response modeling is that past data on consumer and market response to the marketing mix contain valuable information that can enlighten our understanding of response. Those data also enable us to predict how consumers might respond in the future and therefore how best to plan marketing variables (e.g., Tellis & Zufryden, 1995). While no one can assert the future for sure, no one should ignore the past entirely. Thus, we want to capture as much information as we can from the past to make valid inferences and develop good strategies for the future.

Assume that we fit a regression model in which the dependent variable is a brand's sales and the independent variable is advertising or price. Thus,

$$Y_t = \alpha + \beta A_t + \varepsilon_t. \quad (1)$$

Here, Y represents the dependent variable (e.g., sales), A represents advertising, the parameters α and β are coefficients or parameters that the researcher wants to estimate, and the subscript t represents various time periods. A section below discusses the problem of the appropriate time interval, but for now, the researcher may think of time as measured in weeks or days. The ε_t are errors in the estimation of Y_i that we assume to independently and identically follow a normal distribution (IID normal). Equation (1) can be estimated by regression (see Chapter 13). Then the coefficient β of the model captures the effect of advertising on sales. In effect, this coefficient nicely summarizes much that we can learn from the past. It provides a foundation to design strategies for the future. Clearly, the validity, relevance, and usefulness of the parameters depend on how well the models capture past reality. Chapters 13, 14, and 19 describe how to correctly specify those models. This chapter explains how we can implement them in the context of the marketing mix. We focus on advertising and price for three reasons. First, these are the variables most often under the control of managers. Second, the literature has a rich history of models that capture response to these variables. Third, response to these variables has a wealth of interesting patterns or effects. Understanding how to model these response patterns can enlighten the modeling of other marketing variables.

The first step is to understand the variety of patterns by which contemporary markets respond to advertising and pricing. These patterns of response are also called the effects of advertising or pricing. We then present the most important econometric models and discuss how these classic models capture or fail to capture each of these effects.

Patterns of Advertising Response

We can identify seven important patterns of response to advertising. These are the current, shape, competitive, carryover, dynamic, content, and media effects. The first four of these effects are common across price and other marketing variables. The last three are unique to advertising. The next seven subsections describe these effects.

Current Effect

The current effect of advertising is the change in sales caused by an exposure (or pulse or burst) of advertising occurring at the same time period as the exposure. Consider Figure 24.1. It plots time on the *x*-axis, sales on the *y*-axis, and the normal or baseline sales as the dashed line. Then the current effect of advertising is the spike in sales from the baseline given an exposure of advertising (see Figure 24.1A). Decades of research indicate that this effect of advertising is small relative to that of other marketing variables and quite fragile. For example, the current effect of price is 20 times larger than the effect of advertising (Sethuraman & Tellis, 1991; Tellis, 1989). Also, the effect of advertising is so small as to be easily drowned out by the noise in the data. Thus, one of the most important tasks of the researcher is to specify the model very carefully to avoid exaggerating or failing to observe an effect that is known to be fragile (e.g., Tellis & Weiss, 1995).

Carryover Effect

The carryover effect of advertising is that portion of its effect that occurs in time periods following the pulse of advertising. Figure 24.1 shows long (1B) and short (1C) carryover effects. The carryover effect may occur for several reasons, such as delayed exposure to the ad, delayed

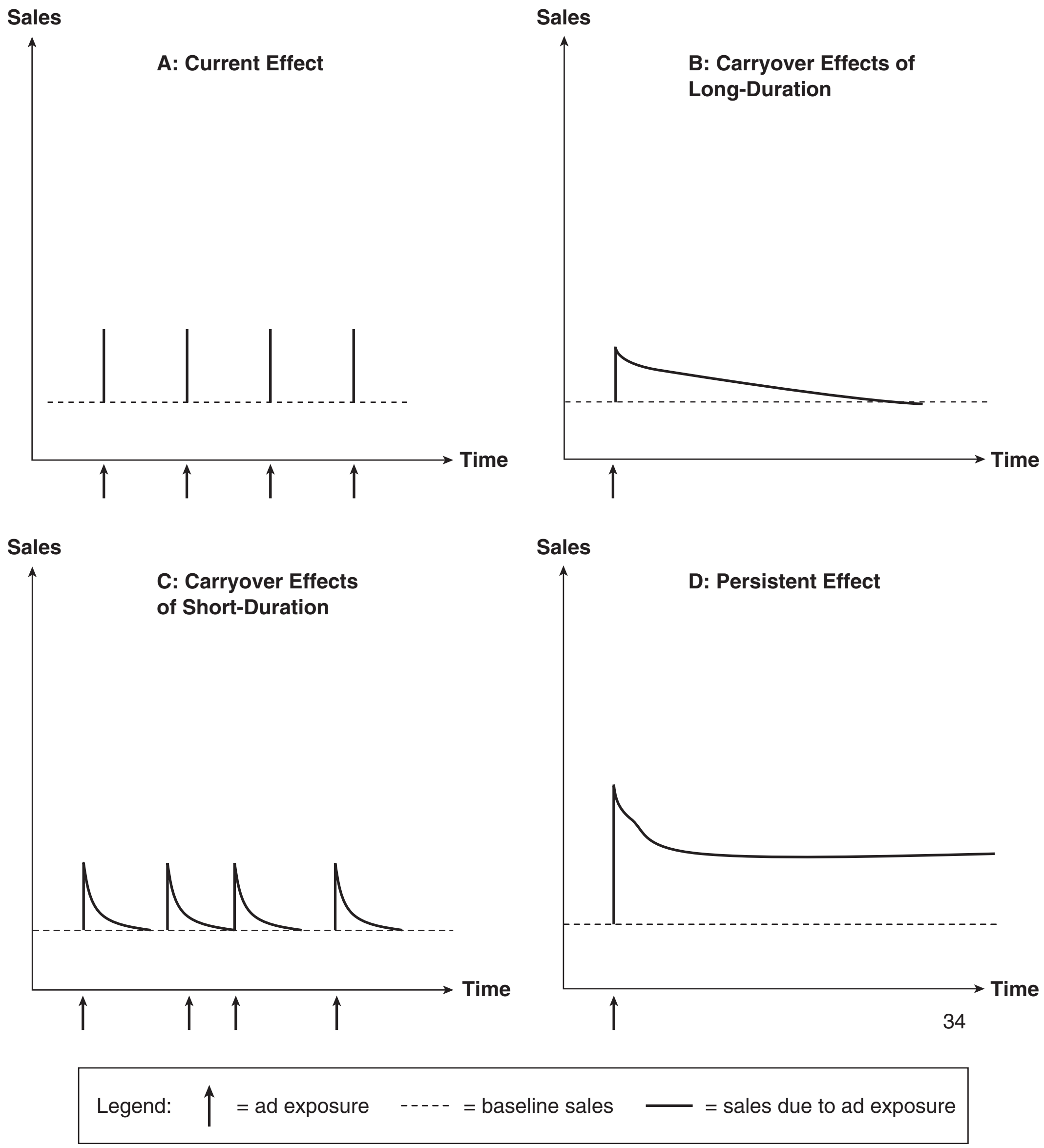

Figure 24.1 Temporal Effects of Advertising

consumer response, delayed purchase due to consumers' backup inventory, delayed purchase due to shortage of retail inventory, and purchases from consumers who have heard from those who first saw the ad (word of mouth). The carryover effect may be as large as or larger than the current effect. Typically, the carryover effect is of short duration, as shown in Figure 24.1C, rather than of long duration, as shown in Figure 24.1B (Tellis, 2004). The long duration that researchers often find is due to the use of data with long intervals that are temporally aggregate (Clarke, 1976). For this reason, researchers should use data that are as temporally disaggregate as they can find (Tellis & Franses, in press). The total effect of advertising from an exposure of advertising is the sum of the current effect and all of the carryover effect due to it.

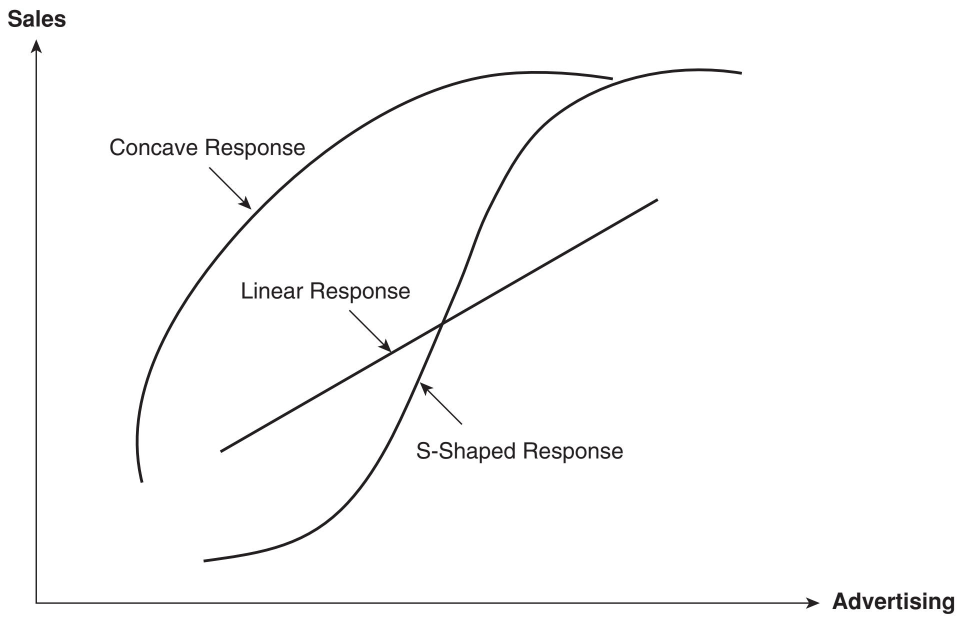

Figure 24.2 Linear and Nonlinear Response to Advertising

Shape Effect

The shape of the effect refers to the change in sales in response to increasing intensity of advertising *in the same time period.* The intensity of advertising could be in the form of exposures per unit time and is also called frequency or weight. Figure 24.2 describes varying shapes of advertising response. Note, first, that the *x*-axis now is the intensity of advertising (in a period), while the *y*-axis is the response of sales (during the same period). With reference to Figure 24.1, Figure 24.2 charts the height of the bar in Figure 24.1A, as we increase the exposures of advertising.

Figure 24.2 shows three typical shapes: linear, concave (increasing at a decreasing rate), and S-shape. Of these three shapes, the S-shape seems the most plausible. The linear shape is implausible because it implies that sales will increase indefinitely up to infinity as advertising increases. The concave shape addresses the implausibility of the linear shape. However, the S-shape seems the most plausible because it suggests that at some very low level, advertising might not be effective at all because it gets drowned out in the noise. At some very high level, it might not increase sales because the market is saturated or consumers suffer from tedium with repetitive advertising.

The responsiveness of sales to advertising is the rate of change in sales as we change advertising. It is captured by the slope of the curve in Figure 24.2 or the coefficient of the model used to estimate the curve. This coefficient is generally represented as β in Equation (1). Just as we expect the advertising sales curve to follow a certain shape, we also expect this responsiveness of sales to advertising to show certain characteristics. First, the estimated response should preferably be in the form of an elasticity. The elasticity of sales to advertising (also called advertising elasticity, in short) is the percentage change in sales for a 1% change in advertising. So defined, an elasticity is units-free and does not depend on the measures of advertising or of sales. Thus, it is a pure measure of advertising responsiveness whose value can be compared across products, firms, markets, and time. Second, the elasticity should neither always increase with the level of advertising nor be always constant but should show an inverted bell-shaped pattern in the level of advertising. The reason is the following.

We would expect responsiveness to be low at low levels of advertising because it would be drowned out by the noise in the market. We would expect responsiveness to be low also at very high levels of advertising because of saturation. Thus, we would expect the maximum responsiveness of sales at moderate levels of advertising. It turns out that when advertising has an S-shaped response with sales, the advertising elasticity would have this inverted bell-shaped response with respect to advertising. So the model that can capture the S-shaped response would also capture advertising elasticity in its theoretically most appealing form.

Competitive Effects

Advertising normally takes place in free markets. Whenever one brand advertises a successful innovation or successfully uses a new advertising form, other brands quickly imitate it. Competitive advertising tends to increase the noise in the market and thus reduce the effectiveness of any one brand's advertising. The competitive effect of a target brand's advertising is its effectiveness relative to that of the other brands in the market. Because most advertising takes place in the presence of competition, trying to understand advertising of a target brand in isolation may be erroneous and lead to biased estimates of the elasticity. The simplest method of capturing advertising response in competition is to measure and model sales and advertising of the target brand relative to all other brands in the market.

In addition to just the noise effect of competitive advertising, a target brand's advertising might differ due to its position in the market or its familiarity with consumers. For example, established or larger brands may generally get more mileage than new or smaller brands from the same level of advertising because of the better name recognition and loyalty of the former. This effect is called differential advertising responsiveness due to brand position or brand familiarity.

Dynamic Effects

Dynamic effects are those effects of advertising that change with time. Included under this term are carryover effects discussed earlier and wearin, wearout, and hysteresis discussed here. To understand wearin and wearout, we need to return to Figure 24.2. Note that for the concave and the S-shaped advertising response, sales increase until they reach some peak as advertising intensity increases. This advertising response can be captured in a static context—say, the first week or the average week of a campaign. However, in reality, this response pattern changes as the campaign progresses.

Wearin is the increase in the response of sales to advertising, from one week to the next of a campaign, even though *advertising occurs at the same level* each week (see Figure 24.3). Figure 24.3 shows time on the *x*-axis (say in weeks) and sales on the *y*-axis. It assumes an advertising campaign of 7 weeks, with one exposure per week at approximately the same time each week. Notice a small spike in sales with each exposure. However, these spikes keep increasing during the first 3 weeks of the campaign, even though the advertising level is the same. That is the phenomenon of wearin. Indeed, if it at all occurs, wearin typically occurs at the start of a campaign. It could occur because repetition of a campaign in subsequent periods enables more people to see the ad, talk about it, think about it, and respond to it than would have done so on the very first period of the campaign.

Wearout is the decline in sales response of sales to advertising from week to week of a campaign, even though advertising occurs at the same level each week. Wearout typically occurs at the end of a campaign because of consumer tedium. Figure 24.3 shows wearout in the last 3 weeks of the campaign.

Hysteresis is the permanent effect of an advertising exposure that persists even after the pulse is withdrawn or the campaign is stopped (see Figure 24.1D). Typically, this effect does not occur more than once. It occurs because an ad established a dramatic and previously unknown fact, linkage, or relationship. Hysteresis is an unusual effect of advertising that is quite rare.

Content Effects

Content effects are the variation in response to advertising due to variation in the content or creative cues of the ad. This is the most

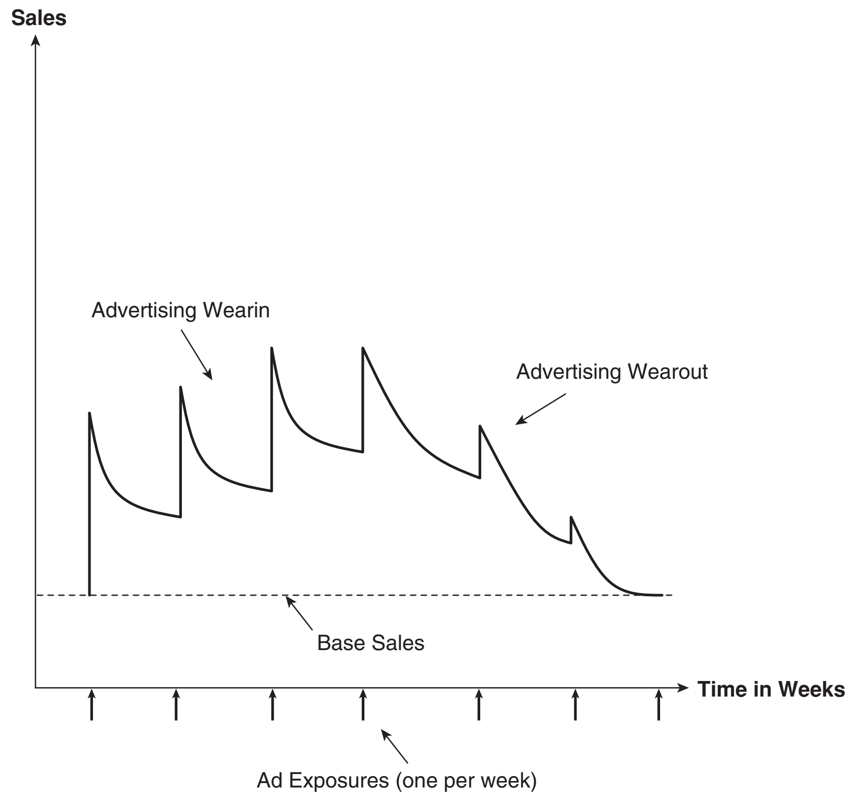

Figure 24.3 Wearin and Wearout in Advertising Effectiveness

important source of variation in advertising responsiveness and the focus of the creative talent in every agency. This topic is essentially studied in the field of consumer behavior using laboratory or theater experiments. However, experimental findings cannot be easily and immediately translated into management practice because they have not been replicated in the field or in real markets. Typically, modelers have captured the response of consumers or markets to advertising measured in the aggregate (in dollars, gross ratings points, or exposures) without regard to advertising content. So the challenge for modelers is to include measures of the *content* of advertising when modeling advertising response in real markets.

Media Effects

Media effects are the differences in advertising response due to various media, such as TV or newspaper, and the programs within them, such as channel for TV or section or story for newspaper.

Modeling Advertising Response

This section discusses five different models of advertising response, which address one or more of the above effects. Some of these models are applications of generic forms presented in Chapters 12, 13, and 14. The models are presented in the order of increasing complexity. By discussing the strengths and weaknesses of each model, the reader will appreciate its value and the progression to more complex models. By combining one or more models below, a researcher may be able to develop a model that can capture many of the effects listed above. However, that task is achieved at the cost of great complexity. Ideally, an advertising model should be rich enough to capture all the seven effects discussed above. No one has proposed a model that has done so, though a few have come close.

Basic Linear Model

The basic linear model can capture the first of the effects described above, the current effect. The model takes the following form:

$$Y_t = \alpha + \beta_1 A_t + \beta_2 P_t + \beta_3 R_t + \beta_4 Q_t + \varepsilon_t. \quad (2)$$

Here, Y represents the dependent variable (e.g., sales), while the other capital letters represent variables of the marketing mix, such as advertising (A), price (P), sales promotion (R), or quality (Q). The parameters α and β_k are coefficients that the researcher wants to estimate. β_k represents the effect of the independent variables on the dependent variable, where the subscript k is an index for the independent variables. The subscript t represents various time periods. A section below discusses the problem of the appropriate time interval, but for now, the researcher may think of time as measured in weeks or days. The ε_t are errors in the estimation of Y_t that we assume to independently and identically follow a normal distribution (IID normal). This assumption means that there is no pattern to the errors so that they constitute just random noise (also called white noise). Our simple model assumes we have multiple observations (over time) for sales, advertising, and the other marketing variables. This model can best be estimated by regression, a simple but powerful statistical tool discussed in Chapter 13. While simple, this model can only capture the first of the seven effects discussed above.

Multiplicative Model

The multiplicative model derives its name from the fact that the independent variables of the marketing mix are multiplied together. Thus,

$$Y_t = \text{Exp}(\alpha) \times A_t^{\beta 1} \times P_t^{\beta 2} \times R_t^{\beta 3} \times Q_t^{\beta 4} \times \varepsilon_t. \quad (3)$$

While this model seems complex, a simple transformation can render it quite simple. In particular, the logarithmic transformation linearizes Equation (3) and renders it similar to Equation (2); thus,

$$\log(Y_t) = \alpha + \beta_1 \log(A_t) + \beta_2 \log(P_t) + \beta_3 \log(R_t) + \beta_4 \log(Q_t) + \varepsilon_t. \quad (4)$$

The main difference between Equation (2) and Equation (4) is that the latter has all variables as the logarithmic transformation of their original state in the former. After this transformation, the error terms in Equation (4) are assumed to be IID normal.

The multiplicative model has many benefits. First, this model implies that the dependent variable is affected by an interaction of the variables of the marketing mix. In other words, the independent variables have a synergistic effect on the dependent variable. In many advertising situations, the variables could indeed interact to have such an impact. For example, higher advertising *combined* with a price drop may enhance sales more than the *sum* of higher advertising or the price drop occurring alone.

Second, Equations (3) and (4) imply that response of sales to any of the independent variables can take on a variety of shapes depending on the value of the coefficient. In other words, the model is flexible enough that it can capture relationships that take a variety of shapes by estimating appropriate values of the response coefficient.

Third, the β coefficients not only estimate the effects of the independent variables on the dependent variables, but they are also elasticities. Estimating response in the form of elasticities has a number of advantages listed above.

However, the multiplicative model has three major limitations. First, it cannot estimate the latter five of the seven effects described above. For this purpose, we have to go to other models. Second, the multiplicative model is unable to capture an S-shaped response of advertising to sales. Third, the multiplicative model implies that the elasticity of sales to advertising is constant. In other words, the percentage rate at which sales increase in response to a percentage increase in advertising is the same whatever the level of sales or advertising. This result is quite implausible. We would expect that percentage increase in sales in response to a percentage increase in advertising would be lower as the firm's sales or advertising become very large. Equation (4) does not allow such variation in the elasticity of sales to advertising.

Exponential Attraction and Multinomial Logit Model

Attraction models are based on the premise that market response is the result of the attractive power of a brand relative to that of other brands with which they compete. The attraction model implies that a brand's share of market sales is a function of its share of total marketing effort; thus,

$$M_i = S_i / \sum_j S_j = F_i / \sum_j F_j, \tag{5}$$

Here, M_i is the market share of the *i*th brand (measured from 0 to 1), S_i is the sales of brand *i*, $\sum_j$ implies a summation of the values of the corresponding variable over all the *j* brands in the market, and F_i is brand *i*'s marketing effort and is the effort expended on the marketing mix (advertising, price, promotion, quality, etc.). Equation (5) has been called Kotler's fundamental theorem of marketing. Also, the right-hand-side term of Equation (5) has been called the attraction of brand *i*. Attraction models intrinsically capture the effects of competition.

A simple but inaccurate form of the attraction model is the use of the relative form of all variables in Equation (2). So for sales, the researcher would use market share. For advertising, he or she would use share of advertising expenditures or share of gross rating points (share of voice) and so on. While such a model would capture the effects of competition, it would suffer from other problems of the linear model, such as linearity in response. Also, it is inaccurate because the right-hand side would not be exactly the share of marketing effort but the sum of the individual shares of effort on each element of the marketing mix.

A modification of the linear attraction model can resolve the problem of linearity in response and the inaccuracy in specifying the right-hand side of the model plus provide a number of other benefits. This modification expresses the market share of the brand as an exponential attraction of the marketing mix; thus,

$$M_i = \text{Exp}\,(V_i) / \sum_j \text{Exp}\, V_j, \tag{6}$$

where M_i is the market share of the *i*th brand (measured from 0 to 1), V_j is the marketing effort of the *j*th brand in the market, $\sum_j$ stands for summation over the *j* brands in the market, Exp stands for exponent, and V_i is the marketing effort of the *i*th brand, expressed as the right-hand side of Equation (2). Thus,

$$V_i = \alpha + \beta_1 A_i + \beta_2 P_i + \beta_3 R_i + \beta_4 Q_i + e_i, \tag{7}$$

where e_i are error terms. By substituting the value of Equation (7) in Equation (6), we get

$$M_i = \text{Exp}\,(V_i)/\sum_j \text{Exp}\, V_j = \text{Exp}(\sum_k \beta_k X_{ik} + e_i)/ \sum_j \text{Exp}(\sum_k \beta_k X_{ik} + e_j), \tag{8}$$

where X_k (0 to m) are the m independent variables or elements of the marketing mix, and $\alpha = \beta_0$ and $X_{i0} = 1$. The use of the ratio of exponents in Equations (6) and (8) ensures that market share is an S-shaped function of share of a brand's marketing effort. As such, it has a number of nice features discussed earlier.

However, Equation (8) also has two limitations. First, it is not easy to interpret because the right-hand side of Equation (8) is in the form of exponents. Second, it is intrinsically nonlinear and difficult to estimate because the denominator of the right-hand side is a sum of the exponent of the marketing effort of each brand summed over each element of the marketing mix. Fortunately, both of these problems can be solved by applying the log-centering transformation to Equation (8) (Cooper & Nakanishi, 1988). After applying this transformation, Equation (8) reduces to

$$\text{Log}(M_i M^-) = \alpha_i^* + \sum_k \beta_k (X_{ik}^*) + e_i^*, \tag{9}$$

where the terms with * are the log-centered version of the normal terms; thus, $\alpha_i^* = \alpha_i - \bar{\alpha}$, $X_{ik}^* = X_{ik} - \bar{X}_i$, $e_i^* = e_i - \bar{e}$, for $k = 1$ to m, and the terms with are the geometric means of the normal variables over the m brand in the market.

The log-centering transformation of Equation (8) reduces it to a type of multinomial logit model in Equation (9). The nice feature of this model is that it is relatively simpler, more easily interpreted, and more easily estimated than Equation (8). The right-hand side of Equation (9) is a linear sum of the transformed independent variables. The left-hand side of Equation (9) is a type of logistic transformation of market share and can be interpreted as the log odds of consumers as a whole preferring the target brand relative to the average brand in the market.

The particular form of the multinominal logit in Equation (9) is aggregate. That is, this form is estimated at the level of market data obtained in the form of market shares of the brand and its share of the marketing effort relative to the other brands in the market. An analogous form of the model can be estimated at the level of an individual consumer's choices (e.g., Tellis, 1988a). This other form of the model estimates how individual consumers choose among rival brands and is called the multinomial logit model of brand choice (Guadagni & Little, 1983). Chapter 14 covers this choice model in more detail than done here.

The multinomial logit model (Equation (9)) has a number of attractive features that render it superior to any of the models discussed above. First, the model takes into account the competitive context, so that predictions of the model are sum and range constrained, just as are the original data. That is, the predictions of the market share of any brand range between 0 and 1, and the sum of the predictions of all the brands in the market equals 1.

Second, and more important, the functional form of Equation (6) (from which Equation (9) is derived) suggests a characteristic S-shaped curve between market share and any of the independent variables (see Figure 24.2). In the case of advertising, for example, this shape implies that response to advertising is low at levels of advertising that are very low or very high. This characteristic is particularly appealing based on advertising theory. The reason is that very low levels of advertising may not be effective because they get lost in the noise of competing messages. Very high levels of advertising may not be effective because of saturation or diminishing returns to scale. If the estimated lower threshold of the S-shaped relationship does not coincide with 0, this indicates that market share maintains some minimal floor level even when marketing effort declines to a zero. We can interpret this minimal floor to be the base loyalty of the brand. Alternatively, we can interpret the level of marketing effort that coincides with the threshold (or first turning point) of the S-shaped curve as the minimum point necessary for consumers or the market to even notice a change in marketing effort.

Third, because of the S-shaped curve of the multinomial logit model, the elasticity of market share to any of the independent variables shows a characteristic bell-shaped relationship with respect to marketing effort. This relationship implies that at very high levels of marketing effort, a 1% increase in marketing effort translates into ever smaller percentage increases in market share. Conversely, at very low levels of marketing effort, a 1% decrease in marketing effort translates into ever smaller percentage decreases in market share. Thus, market share is most responsive to marketing effort at some intermediate level of market share. This pattern is what we would expect intuitively of the relationships between market share and marketing effort.

Despite its many attractions, the exponential attraction or multinomial model as defined above does not capture the latter four of the seven effects identified above.

Koyck and Distributed Lag Models

The Koyck model may be considered a simple augmentation of the basic linear model (Equation (2)), which includes the lagged dependent variable as an independent variable. What this specification means is that sales depend on sales of the prior period and all the independent variables that caused prior sales, plus the current values of the same independent variables.

$$Y_t = \alpha + \lambda Y_{t-1} + \beta_1 A_t + \beta_2 P_t + \beta_3 R_t + \beta_4 Q_t + \varepsilon_t. \quad (10)$$

In this model, the current effect of advertising is β_1, and the carryover effect of advertising is $\beta_1\lambda/(1-\lambda)$. The higher the value of λ, the longer the effect of advertising. The smaller the value of λ, the shorter the effects of advertising, so that sales depend more on only current advertising. The total effect of advertising is $\beta_1/(1-\lambda)$.

While this model looks relatively simple and has some very nice features, its mathematics can be quite complex (Clarke, 1976). Moreover, readers should keep in mind the following limitations of the model. First, this model can capture carryover effects that only decay smoothly and do not have a hump or a nonmonotonic decay. Second, estimating the carryover of any one variable is quite difficult when there are multiple independent variables, each with its own carryover effect. Third, the level of data aggregation is critical. The estimated duration of the carryover increases or is biased upwards as the level of aggregation increases. A recent paper has proved that the optimal data interval that does not lead to any bias is not the interpurchase time of the category, as commonly believed, but the largest period with at most one exposure and, if it occurs, does so at the same time each period (Tellis & Franses, in press).

The distributed lag model is a model with multiple lagged values of both the dependent variable and the independent variable. Thus,

$$\begin{aligned} Y_t = \alpha + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \lambda_3 Y_{t-3} + \ldots \\ + \beta_{10} A_t + \beta_{11} A_{t-1} + \beta_{12} A_{t-2} + \ldots \\ + \beta_2 P_t + \beta_3 R_t + \beta_4 Q_t + \varepsilon_t. \end{aligned} \tag{11}$$

This model is very general and can capture a whole range of carryover effects. Indeed, the Koyck model can be considered a special case of distributed lag model with only one lagged value of the dependent variable. The distributed lag model overcomes two of the problems with the Koyck model. First, it allows for decay functions, which are nonmonotonic or humped shaped (see Figure 24.4). Second, it can partly separate out the carryover effects of different independent variables. However, it also suffers from two limitations. First, there is considerable multicollinearity between lagged and current values of the same variables. Second, because of this problem, estimating how many lagged variables are necessary is difficult and unreliable. Thus, if the researcher has sufficient extensive data that minimize the latter two problems, then he or she should use the distributed lagged model. Otherwise, the Koyck model would be a reasonable approximation.

Hierarchical Models

The remaining effects of advertising that we need to capture (content, media, wearin, and wearout) involve changes in the responsiveness itself of advertising (i.e., the β coefficient) due to advertising content, media used, or time of a campaign. These effects can be captured in one of two ways: dummy variable regression or a hierarchical model.

Dummy variable regression is the use of various interaction terms to capture how advertising responsiveness varies by content, media, wearin, or wearout. We illustrate it in the context of a campaign with a few ads. First, suppose the advertising campaign uses only a few different types of ads (say, two). Also, assume we start with the simple regression model of Equation (3). Then we can capture the effects of these different ads by including suitable dummy variables. One simple form is to include a dummy variable for the second ad, plus an interaction effect of advertising times this dummy variable. Thus,

$$\begin{aligned} Y_t = \alpha + \beta_1 A_t + \delta A_t A_{2t} + \\ \beta_2 P_t + \beta_3 R_t + \beta_4 Q_t + \varepsilon_t, \end{aligned} \tag{12}$$

where A_{2t} is a dummy variable that takes on the value of 0 if the first ad is used at time t and the value of 1 if the second ad is used at time t. δ is the effect of the interaction term ($A_t A_{2t}$). In this case, the main coefficient of advertising, β_1, captures the effect of the first ad, while the coefficients of β_1 plus that of the interaction term (δ) capture the effect of the second ad. While simple, these models quickly become quite complex when we have multiple ads, media, and time periods, especially if these are occurring simultaneously. This is the situation in real markets. The problem can be solved by the use of hierarchical models.

Hierarchical models are multistage models in which coefficients (of advertising) estimated in one stage become the dependent variable in the other stage. The second stage contains the characteristics by which advertising is likely to vary in the first stage, such as ad content, medium, or campaign duration. Consider the following example.

Example

A researcher gathers data about the effect of advertising on sales for a brand of one firm over a 2-year period. The firm advertises the brand using a large number of different ads (or copy content), in campaigns of varying duration (say, 2 to 8 weeks), in a number of different cities or

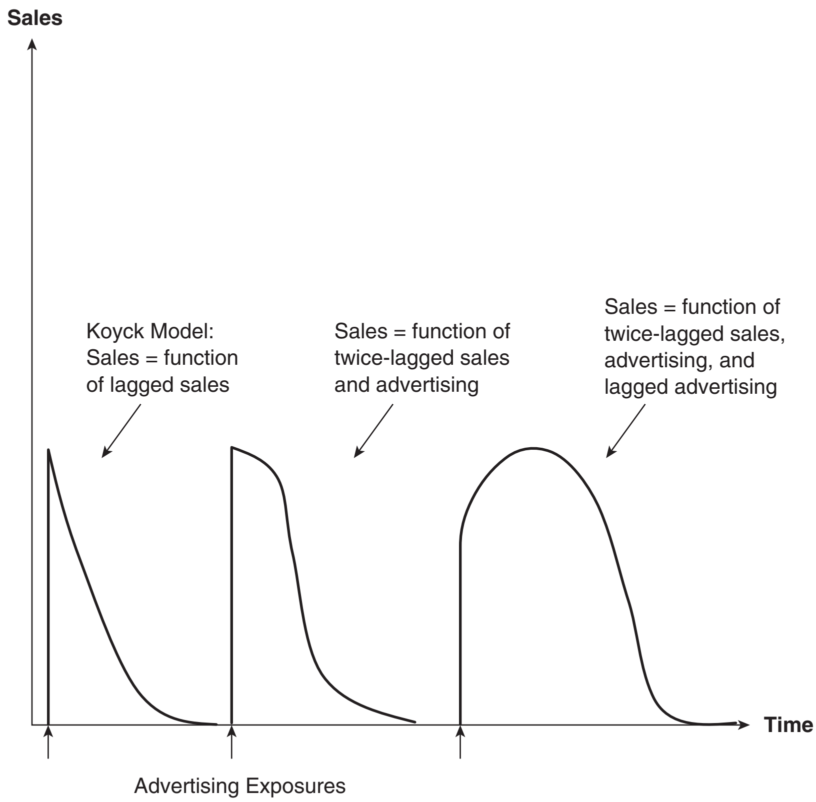

Figure 24.4 Alternate Shapes of Advertising Carryover

markets. Assume that the researcher has data at a highly disaggregate level, say the hour of the day. Such data are possible because of electronic databases such as that recorded by Internet retailers, telemarketers, or retail firms with scanners. The researcher analyzes the effect of advertising on sales, separately for each city, campaign (ad), and week of the campaign. These effects vary substantially across the various estimations of the model. Why do they vary?

The researcher suspects that the variation could be due to varying responsiveness in markets, or by campaign, or by week of the campaign. The researcher has information on all these three factors (market, campaign, and week of campaign). Then in a second-stage model, the researcher can analyze how the coefficients of advertising estimated in the first stage vary due to these three factors. The dependent variable is the coefficients of advertising from the first stage, and the independent variables are the factors that gave rise to that coefficient. Such a multistage model is called a hierarchical model (e.g., Chandy, Tellis, MacInnis, & Thaivanich, 2001; Tellis, Chandy, & Thaivanich, 2000).

Two features are essential for hierarchical models. First, we should be able to obtain multiple estimates of the effects (or coefficient values) of advertising on some dependent variable such as sales or market share for the same brand across different contexts such as at least one of the following: the ad campaign, week of the campaign, market, or medium. Then we can use the estimates of the effects of advertising from the first stage as dependent variables in the second stage. Second, as far as possible, we need to minimize excessive covariation among factors. Thus, a particular ad should not always occur with a particular channel, or an ad of a particular duration should not always be run in a particular channel. Such co-occurrence leads to the problem of multicollinearity among the created variables in the second-stage model (see Chapter 13). As long as the three factors have sufficient cross-variation, estimates of the second-stage model should be reliable.

Depending on the richness of the data, hierarchical models can estimate the last three effects of advertising that we identified above. That is, with such models and given suitable data, the researcher can estimate what ad content is the most effective, what duration of the campaign is the most effective, and which media are the most effective. The duration of the campaign could be estimated in terms of weeks. For example, if the effectiveness of the ad first increases slowly and then decreases suddenly, one could conclude that wearin is slow but wearout is rapid. On the other hand, if the effectiveness of the ad steadily declines over time, then there is no wearin, and wearout sets out from the start. Furthermore, if the data are sufficiently rich and detailed, the researcher can also obtain interaction effects such as which media are most suitable for particular ads or which ad content needs to be run over campaigns of long versus short duration.

Note that to address all of the seven effects of advertising identified above, the researcher would have to use a hierarchical model, which itself contains an exponential attraction or multinomial logit model with a Koyck-type or distributed lag enhancement. In other words, suitably integrating models described above would enable a researcher to address the most important phenomena associated with advertising. In reality, such fully integrated models that can capture all the effects of advertising are very complex and require substantial data (e.g., see Chandy et al., 2001). If researchers want to focus on only a few effects or their data are not rich, they might want to simplify the model they use to focus on only the most important effects.

Patterns and Models of Price Response

The first four effects of advertising response also apply to price: current, shape, competition, and carryover effects. The current effect of price is the changes in sales that occur in the same period as that in which prices change. In contrast to response to advertising, response to price is typically strong and immediate, with most of the effect lasting in the current period (Sethuraman & Tellis, 1991).

However, price changes can also have carryover effects. These effects could occur because consumers take time to learn of the price change, wait to respond until their next shopping trip, or wait to respond because of their current inventory. Typically, carryover effects are less pronounced for price than for advertising. One type of carryover is the negative sales following a price cut, because consumers buy excess stock during the discount and then hold back regular purchases until they deplete their stocks.

The exponential attraction or multinomial logit model specified for advertising response also serves very well to capture S-shaped response and competitive effects, if any, in response to pricing. In addition, the integration of these models with a Koyck or distributed lag specification can capture any carryover effects that may exist in response to pricing.

In addition, response to price has three more effects that are unique to price: promotional price effect, reference price effect, and price interaction effect. To capture the three effects, the researcher has merely to modify the linear, multinomial logit, or distributed lag model by including relevant independent variables. The basic structure of the model need not change. Thus, in the interests of parsimony, here we discuss only the unique effects of pricing and how modifications of the classic models discussed above can capture these effects.

Modeling Promotion Price Effect

A pervasive feature of pricing in contemporary markets is that prices are constantly in flux. Retailers have a certain list price, and frequently for a variety of reasons, they offer discounts or "sales" from these prices (Tellis, 1986). Thus, pricing strategies have two components: (1) a list price component that is basically how a brand is listed on price relative to other brands and (2) a promotion price component, which basically involves a temporary discount off this list price. So, models of response to pricing should contain both of these components to correctly specify and fully capture all the effects of price.

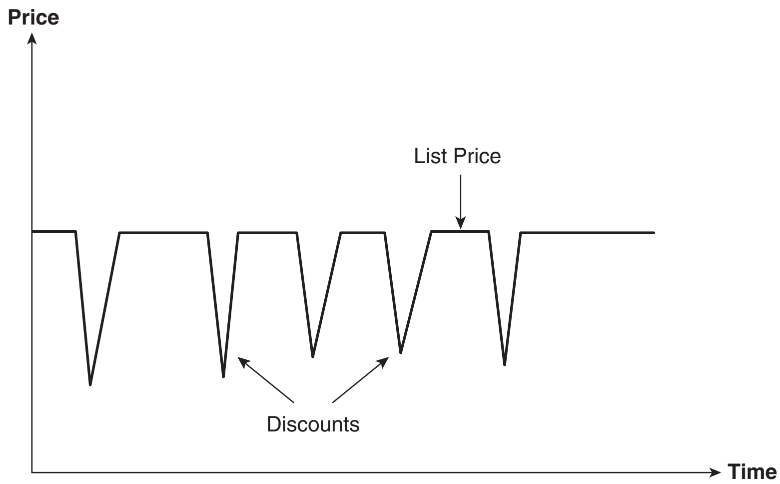

Figure 24.5 Price Path of One Brand in One Store Over Time

Assume one has chosen the multinominal logit model discussed above. Then, to fully capture the promotional price effect, one would use two independent variables for price, instead of only one: One variable would represent the list price of the brand; the other variable would represent the promotion price of the brand. The key question would be how to measure the list and promotion price.

In markets today, firms generally keep the list price of the brand high for an extended period of time but occasionally drop its price by offering a sales or discount (Tellis, 1998). Thus, one can define and capture the list price as the high modal price of the brand over a given time horizon (see Figure 24.5). One can define the promotional price or discount of the brand as the list price minus the actual price charged or paid in a particular time period within that horizon. One would use the same rules to compute the list and promotional prices for competing brands.

The estimated coefficients (elasticities) of these variables would then reflect the response of markets to these respective variables. By their definition, the effect of the list price would generally be negative. That is, the higher the list price of the brand, the lower its sales or market share. The effect of the promotional price would generally be positive. That is, the steeper the promotional discount of the brand, the higher its sales or market share.

Modeling Reference Price Effect

Reference prices are latent internal norms that consumers use as a basis against which to compare current prices (Tellis, 1998; Winer, 1986). Reference prices are not observed and cannot be ascertained by survey because of the problem of demand bias. Even if they did not exist, consumers would be tempted to answer in the affirmative about them just to please the researcher. The best way to test for reference prices is by the prediction of behavior with and without reference prices. For example, a researcher can ascertain a model's improvement in fit with the data, if any, from the inclusion of terms that capture reference price.

Current research suggests at least two components of reference price (Rajendran & Tellis, 1994): first, a temporal or internal reference price based on memory that probably develops in response to past prices a consumer has paid and, second, an external or contextual reference price based on visible prices that probably relates to the prices of other competing brands available to the consumer at the time of purchase. A complete model of response to pricing should capture these effects of reference price. Any of the models discussed above can account for reference price effects by including independent variables for these effects. In effect, instead of a single variable for price, the researcher would include a variable for the temporal reference price minus price paid, plus another variable for the contextual reference price minus price paid. The next problem is how exactly to measure these reference prices.

To measure the contextual reference price of a target brand, the researcher could use either one of the average of the other brands' prices or the lowest among the other brands' prices or the price of the leading rival brand (Rajendran & Tellis, 1994). The other or rival brands being considered in this case are those with which the target brand is available. Which of these three prices a researcher uses depends on which price is most salient to consumers when they make decisions based on price. In the absence of a strong theory about this issue, a researcher would try out each of these three reference prices and use the one that gives the best fit with the data.

To capture the temporal reference price, the researcher would use some moving average of past prices that the consumer has used for the target brand. Instead of a simple average, some researchers advocate a weighted moving average of past prices. The key issue here is, how does one estimate the weights and the numbers of prior periods that should be included in the definition? The current thinking is that one should fit a time-series model that best captures the string of past prices for a brand (Winer, 1986). The logic for this thinking is that the prices that can best be predicted are those that a consumer is mostly likely to be able to recollect and respond to. However, there is no absolute rule that any one measure of past prices is the best for the temporal reference price component. In effect, a researcher would use that component that he or she finds to fit the data best.

Modeling Promotional and Reference Price Effects Jointly

A model can get quite unwieldy if one attempts to capture both promotional and reference price effects and, for each of these, capture both temporal and contextual components. Fortunately, reference price effects are probably related to promotional price effects. In particular, list prices are more likely to need a contextual or external reference price. The reason is that list prices do not change much over time, so consumers probably form them from the list prices of competing brands at the point of purchase. On the other hand, promotional prices are more likely to be compared to a temporal or internal reference price because they vary over time and depend on consumer memory and experience of these prices.

Thus, despite many pricing effects, a researcher might capture most of these effects parsimoniously with just two independent variables for price. The first variable would be the reference list price minus the actual list price. This term would capture the effect of list prices relative to contextual reference prices. The second term would be the temporal reference discount minus the discount actually obtained. This term would capture the effect of discounts with regard to temporal reference prices. The discount itself is the list price minus the actual price paid at any one period.

Modeling Interaction Effects

Often, marketing variables affect consumers synergistically. That is, the effect of two of them together is greater than the sum of the effect of each of them separately. We refer to this synergistic effect as an interaction effect. One might argue that the whole concept of the marketing mix is that these variables do not act alone but have some joint effect that is much greater than the sum of the parts. The general way in which response models capture interaction effects is by including an additional term that is formed by the product of the two variables that interact. For example, if the researcher believes that advertising would be more effective during the time of a discount, the researcher would include a new independent variable formed from the multiplication of advertising and discounts.

When one already has a large number of independent variables, some of which have multiple components (such as lagged values of advertising or temporal and contextual reference prices), then testing out all sorts of interactions can get quite complex. What is needed is a model that can do so parsimoniously. Some of the past models may do so under certain assumptions.

Consider the multiplicative model in Equation (3). This model in its original form (with all the variables measured naturally) implies that sales result from the multiplicative mix of the independent variables. In other words, it assumes that sales result from the interaction of the marketing mix. However, in its logarithmic form (after taking logs of all the variables) in Equation (4), which is used to linearize and estimate the model, it no longer contains interaction effects. So if theory suggests that the interaction effects hold in the natural state of the variables but not in their logarithmic state, then the multiplicative model serves as a parsimonious means of capturing those interaction effect. Alternatively, if the researcher believes that the interaction effects persist even after taking logs of the natural variables or if the researcher is not sure, he or she could just run a model that includes additional interaction terms of the log of the marketing variables suspected to have interaction effects.

If the researcher has reason to believe that a strong interaction effect exists between some variables and the researcher is using a model other than the multiplicative model, then he or she is best advised to model the interaction effect explicitly. This modeling can be achieved by including an additional independent variable formed by multiplication of those variables that the researcher assumes do interact with each other.

A Partially Integrated Hierarchical Model for Ad Response

No researcher has published a model that captures all of the seven characteristics of marketing-mix models. However, a recent example published in two studies by a team of four authors shows how one could capture all of these effects except competition. Now, many readers will argue that competition is pervasive in markets today and is the most important dimension to capture. However, in this particular example, competitors were not present. Also, advertising was the only element of the marketing mix that the firm used. Given these two caveats, the authors were able to integrate the other six desirable characteristics of marketing models quite nicely.

This example is due to a study done by Tellis et al. (2000) and Chandy et al. (2001). The researchers have referrals (sales) and TV advertising data for a referral service over several years across more than 30 cities. In each city, the service provider can draw from a bank of about 70 creatives developed over the years. Fortunately, the firm uses different creatives in different cities, in each of which the firm has operated for a varying length of time. The researchers were able to describe the differences in those creatives by a set of key characteristics, such as the use of emotion, argument, endorser, certain types of copy, and so on. They were also able to calibrate differences in the various cities by the age of the market at which time the ad was aired.

Given this scenario, a first-stage model could explain what effects each creative has in each city. Then, a second-stage model can explain how those effects vary by type of creatives and type of city. This is a hierarchical model. We now proceed to describe the equations in each stage.

Stage 1: Estimating Response to Ads (Creatives)

The authors began with a distributed lag model such as that in Equation (11). The authors then included a dummy variable in the model for the presence or absence of each creative. The coefficient of this variable determines how the effect of advertising varies from the common effect captured in Equation (11) due to the use of a particular creative and the age of the market at that time. The authors also included many control variables to account for other differences, such as hour of the day and day of the week when the sales occurred, the station and day-part (morning or evening) in which the ad aired, and whether the service was open.

The first-stage model is

$$R = \alpha + (\mathbf{R}_{-l}\lambda + \mathbf{A}\beta_A + \mathbf{C}\beta_c + \mathbf{S}\beta_S + \mathbf{SH}\beta_{SH} + \mathbf{HD}\beta_{HD} + \mathbf{A}_M\beta_M)\, O + \varepsilon_r, \qquad (13)$$

where

R = a vector of referrals by hour,

$\mathbf{R}_{-l}$ = a matrix of lagged referrals by hour,

$\mathbf{A}$ = a matrix of current and lagged ads by hour,

$\mathbf{C}$ = a matrix of dummy variables indicating whether a creative is used in each hour,[1]

$\mathbf{S}$ = a matrix of current and lagged ads in each TV station by hour,

$\mathbf{A}_M$ = a matrix of current and lagged morning ads by hour,

H = a matrix of dummy variables for time of day by hour,

D = a matrix of dummy variables for day of week by hour,

O = a vector of dummies recording whether the service is open by hour,

α = constant term to be estimated,

λ = a vector of coefficients to be estimated for lagged referrals,

β_i = vectors of coefficients to be estimated, and

ε_t = a vector of error terms initially assumed to be IID normal.

Note that in this study, the authors are able to capture many of the key effects of advertising. For example, β_A captures the main effect of advertising by hour of the day. A combination of λ and β_A captures the carryover effect of advertising. β_c captures the effects of various creatives that were used, plus the main effects of advertising by hour of the day. β_S captures the effect of the various media (TV stations) that were used.

Note that the authors included the creatives as dummy variables in Equation (13), indicating whether a creative is used in a particular market. They chose to drop the creatives that had an average effectiveness and to include only those that were significantly above or below the average. Thus, the coefficient of a creative in Equation (13) represents the increase or decrease in expected referrals due to that creative, relative to the average of creatives in that particular market. This specification had the most practical relevance. Managers are not interested much in a global optimization of the best mix of creatives. Rather, they are interested in making improvements over their strategy in the previous year. For this reason, they seek analyses that highlight the best creatives (to use more often) or the worst creatives (to drop).

The results showed that although advertising has small effects, these effects varied dramatically by type of ad and TV channel. Thus, managers could drop the least effective ads and TV channels and spend more on the most effective ads and TV channels. The detailed data and specification of the model revealed a number of other interesting phenomena about how advertising's effects vary and decay by time of the day and day of the week.

Stage 2: Explaining Effectiveness of Ad Response by Type of Creative

In the second stage, the authors collected the coefficients (β_c) for each creative for each market (m) in which it is used and explained their variation as a function of creative characteristics and the age of the market in which it ran as follows:

$$\begin{aligned}\beta_{c,m} = {} & \varphi_1 \text{ Argument}_c + \varphi_2 (\text{Argument}_c \times \text{Age}_m) \\ & + \varphi_3 \text{Emotion}_c + \varphi_4 (\text{Emotion}_c \times \text{Age}_m) \\ & + \varphi_5 \text{800 Visible}_c + \varphi_6 (\text{800 Visible}_c \\ & \times \text{Age}_m) + \varphi_7 \text{Negative}_c + \varphi_8 (\text{Negative}_c \\ & \times \text{Age}_m) + \varphi_9 \text{Positive}_c + \varphi_{10} \\ & (\text{Positive}_c \times \text{Age}_m) + \varphi_{11} \text{Expert}_c + \varphi_{12} \\ & (\text{Expert}_c \times \text{Age}_m) + \varphi_{13} \text{Nonexpert}_c + \varphi_{14} \\ & (\text{Nonexpert}_c \times \text{Age}_m) + \varphi_{15} \text{Age}_m + \varphi_{16} \\ & (\text{Age}_m)^2 + \boldsymbol{\Gamma}\,\textbf{Market} + v, \end{aligned} \qquad (14)$$

where

$\beta_{c,m}$ = coefficients of creative c in market m from Equation (13),

Age = market age (number of weeks since the inception of service in the market),

Market = matrix of market dummies,

$\boldsymbol{\Gamma}$ **=** vector of market coefficients,

v = vector of errors,

φ = second-stage coefficients to be estimated, and other variables are as defined in Equation (4).

The characteristics of creatives that were particularly important were the use of argument, emotion, expert endorsers, visibility of the brand name, negative versus positive arguments, and expert versus nonexpert endorsers. The authors' most important finding was that emotional appeals were effective in mature markets while argument appeals were effective in new markets. Furthermore, a nonlinear regression of the effectiveness of ads on the age of the creatives enabled the researchers to assess the effects of wearin and wearout. They found that ads have no wearin period, and wearout starts from the very first week of the campaign and is steepest in the first few weeks. Thus, frequently changing campaigns and developing new campaigns would be very useful. When developing new campaigns, using appeals that were the most effective for the age of the market would be highly advisable.

Conclusion

Planning the marketing mix is a central task in marketing management. Prudent planning requires that marketing managers take into account how markets have responded to the marketing mix in the past. The underlying assumption is not that the past predicts the future with certainty but that it contains valuable lessons that might enlighten the future.

The econometrics of response modeling describes how a researcher should model response to the marketing mix so as to capture the most important effects validly. This chapter provides an overview of the essential issues and principles in this area. It first describes the important effects that occur in markets today. It then discusses the strengths and limitations of various models that capture those effects.

The chapter focuses on two elements of the marketing mix: advertising and pricing. This focus is because the variables are the most commonly managed and analyzed and encompass a wide range of response patterns. Understanding how to model response to these two variables should provide researchers with the essential tools to model response to other elements of the marketing mix. The chapter provides references to articles and chapters of this book that provide further details on these issues.

Note

1. We use *C* to refer to the matrix of creatives here and *c* to refer to individual creatives later in the chapter.

References

Chandy, R., Tellis, G. J., MacInnis, D., & Thaivanich, P. (2001). What to say when: Advertising appeals in evolving markets. *Journal of Marketing Research, 38,* 399–414.

Clarke, D. G. (1976). Econometric measurement of the duration of advertising effect on sales. *Journal of Marketing Research, 13,* 345–357.

Cooper, L. G., & Nakanishi, M. (1988). *Market share analysis.* Norwell, MA: Kluwer.

Guadagni, P., & Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2,* 203–238.

McCarthy, J. (1996). *Basic marketing: A managerial approach* (12th ed.). Homewood, IL: Irwin.

Rajendran, K. N., & Tellis, G. J. (1994). Is reference price based on context or experience? An analysis of consumers' brand choices. *Journal of Marketing*, *58,* 10–22.

Sethuraman, R., & Tellis, G. J. (1991). An analysis of the tradeoff between advertising and pricing. *Journal of Marketing Research, 31,* 160–174.

Tellis, G. J. (1986). Beyond the many faces of price: An integration of pricing strategies. *Journal of Marketing, 50,* 146–160.

Tellis, G. J. (1988a). Advertising exposure, loyalty and brand purchase: A two-stage model of choice. *Journal of Marketing Research, 15,* 134–144.

Tellis, G. J. (1988b). The price sensitivity of competitive demand: A meta-analysis of sales response models. *Journal of Marketing Research, 15,* 331–341.

Tellis, G. J. (1989). Interpreting advertising and price elasticities. *Journal of Advertising Research, 29*(4), 40–43.

Tellis, G. J. (1998). *Advertising and sales* promotion *strategy.* Reading, MA: Addison-Wesley.

Tellis, G. J. (2004). *Effective advertising: How, when, and why advertising works.* Thousand Oaks, CA: Sage.

Tellis, G. J., Chandy, R., & Thaivanich, P. (2000). Decomposing the effects of direct advertising: Which brand works, when, where, and how long? *Journal of Marketing Research, 37,* 32–46.

Tellis, G. J., & Franses, P. H. (in press). The optimal data interval for econometric models of advertising. *Marketing Science.*

Tellis, G. J., & Weiss, D. (1995). Does TV advertising really affect sales? *Journal of Advertising, 24*(3), 1–12.

Tellis, G. J., & Zufryden, F. (1995). Cracking the retailer's decision problem: Which brand to discount, how much, when and why? *Marketing Science, 14*(3), 271–299.

Winer, R. (1986). A reference price model for demand of frequently purchased goods. *Journal of Consumer Research, 13,* 250–256.