Operations Research

Feature Article—Aggregate Advertising Models: The State of the Art

John D. C. Little,

Feature Article

Aggregate Advertising Models: The State of the Art

JOHN D. C. LITTLE

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received January 1979; accepted April 1979)

Aggregate advertising models relate product sales to advertising spending for a market as a whole Although many models have been built, they frequently contradict each other and considerable doubt exists as to which models best represent advertising processes. An increasingly rich literature of empirical studies helps resolve these issues by revealing major advertising phenomena that models should encompass. These include sales responding upward and downward at different rates, steady state response that can be concave or S-shaped and can have positive sales at zero advertising, sales affected by competitive advertising; and advertising dollar effectiveness that can change over time

A review of aggregate models developed on a priori grounds brings out similarities and differences among those of Vidale and Wolfe, Nerlove and Arrow, Little, and others and identifies ways in which the models agree or disagree with observed phenomena A Lanchester-motivated structure generalizes many features of these models and conforms to some but not all of the empirical observations. Although econometric studies have revealed important empirical insights, the most frequently used structural forms do not model certain key phenomena, most notably different rise and decay rates Future work must join better models with more powerful calibration methods

IGNORANCE of advertising response phenomena, inability to make good measurements, and lack of a theory to organize existing knowledge contribute to great waste in advertising. Contradictions abound. For example, advertising partisans in one company declare that certain markets should receive more advertising because "the brand is strong there and we should take advantage of its momentum." Then, a few minutes later, the same people propose that other markets should also receive more because "industry sales are strong there and our share is low," which, freely translated, means "the brand is weak there and we don't have any momentum." One often sees media scheduled in intensive "flights" so that "the message can be heard through the noise," but, if someone asks why not make the flight half as long and twice as intense,

Operations Research
Vol 27, No 4, July-August 1979

0030-364X/79/2704-0629 $01 25

or else twice as long and half as intense, no good answer can be given.

In one company the brand managers push to spend their budgets early in the calendar year. Is this because of product seasonality? Or a belief in the effectiveness of campaigns lasting six months? No, it is because corporate management has a reputation for calling back unspent monies to improve earnings in the fourth quarter. Brand management responds by spending all its money in the spring. One might suspect that management in this company is not quite sure what it is getting for its advertising dollars. In most companies, advertising strategy is subject to intermittent upheavals. Sometimes this happens brand by brand—each year one or two products undergo an agonizing reappraisal. At other times a whole division will go through a convulsion. Perhaps these strategy shifts are appropriate, but rarely is there any clear reason why the reexamination should be taking place for one brand and not another.

After a substantial change, marketing management watches sales carefully and, more often than not, expresses satisfaction. Yet, though a major strategy shift offers a unique opportunity for measurement (say, by holding out some control markets), such steps are virtually never taken.

Advertising also is full of fads. Clearly a company's ads are conspicuous. (They had better be!) Everybody from the president's wife to the newest clerk voices an opinion. Clever copy becomes a conversation piece overnight. ("We try harder," "I can't believe I ate the whole thing.") Innovations perceived as successful are quickly imitated by others, rightly or wrongly. Low key testimonials, comparison advertising and humor have been up and down over the past few years. Mature authority figures seem to be undergoing a revival at the moment. It is an exciting world of good showmanship where strategy changes are conceived, packaged and sold with many of the appeals that characterize advertising itself.

And, to a great extent, this is as it should be. Good strategy requires imagination and style and always will. At the same time, strategy emerges best from a foundation of reliable facts and sound analysis. These are not easy to come by.

The management science/operations research fraternity has nibbled at advertising issues. Moderate heartburn has been a fairly common result. Yet, there have certainly been successes, one or two of which have been widely publicized. See, for example, Weinberg [72], Rao [52] and Ackoff and Emshoff [1]. Other workers have often found these studies hard to duplicate, perhaps because marketing situations differ from company to company or, more likely, because studies to date simply do not supply enough knowledge to provide an adequate foundation for imitation. Quantitative understanding of advertising processes has made some headway but the job is far from done and the available material needs pulling together. This paper takes on part of that job by examining aggregate response models.

A basic OR/MS goal is to find good models. But what is a good model? It depends. We should tailor a model to fit the job at hand. Lilien [32] calls this "model relativism." Urban [70] expresses the same thought when he says the model builder should state the purposes of his model in advance. All right, we want advertising response models that will be useful for (1) tracking and evaluating advertising performance, (2) diagnosing market changes and (3) incorporation into decision models. Although we shall not address decision models per se, they should contain response models with the necessary phenomena to assist meaningfully on (1) annual budgets, (2) geographic allocation and (3) allocation over time. Two other important areas are media and copy. These enter our discussion but will not be treated with the detail required for incorporation into decision models.

In focusing on the response model rather than the decision model, we differ from the many writers who seek to characterize optimal policies once the response model is given. For an extensive review of this literature, see Sethi [59].

Attainment of our goals requires dynamic models that relate advertising spending to sales. We confine attention to established products since they blot up most of the money and since new products use special models. We focus on macro- or aggregate models rather than models of individual customer behavior for two reasons. First, most micromodels so far have been thin on either empirical data or marketing control variables (especially advertising) or both. Second, the most convincing data sources available to companies for calibrating advertising models today are aggregate in nature (historical time series at a national or market level and field experiments). This is not to play down the importance of modeling individual customer response to advertising (see, for example, the media selection models of Little and Lodish [35], Gensch [16], Zufryden [75] and Starr [65]). Rather it is to say that the catalog of advertising effects presented here comes almost entirely from aggregate data and so is inadequate to resolve most micromodeling questions. We note, however, that micromodels will have to reproduce the empirical macroeffects reported here.

1. CONTROVERSIES, CONFUSIONS, AND CONTRADICTIONS

The advertising models in the OR/MS literature are not especially consistent with each other nor with such measurements and data as are available. We identify three areas of controversy: shape, dynamics, and interactions.

Shape. By shape we mean the shape of a curve showing sales response to advertising under steady state conditions. In other words, if a set of different advertising rates were tested with other market influences held fixed, and brand sales were measured each time after the market came to equilibrium, what would a plot of sales rate vs. advertising rate look like? *Is such a relationship linear*? Many econometric analyses implicitly assume it to be. *What are sales with zero advertising*? A good many theoretical models imply sales would be zero. *Is response S-shaped*? Most existing models do not permit such a possibility, and yet many media schedules contain "flights" whose justification seems to be based on belief in a threshold or S-shape in the curve. *Do large amounts of advertising depress sales*? So claim some writers but few models accommodate it.

Dynamics. How fast do sales respond when advertising is increased? In the process of calibrating marketing models, the author has often asked marketing managers the following question. What percent of the long run response to an advertising increase would you expect to obtain in the first year? A typical answer would be 60% and the range might run from 30% to 80%. It is interesting to compare these values with the data in the next section.

How fast do sales decay when advertising is decreased? Strong marketing men turn pale when advertising cuts are proposed, even if only for test purposes. "We might lose the brand franchise," they say. Their pallor may be role-playing because companies under financial stress regularly cut budgets drastically, apparently believing that the brand will survive.

Still another question is: *Does hysteresis ever exist*? In other words, are there circumstances under which sales would increase with increased advertising and stay there after withdrawal of advertising? Or, in the opposite direction, could a competitor take away sales and share by increasing advertising, and the brand find it difficult to regain position? Very few marketing models exhibit such a phenomenon, but some people believe it to exist in practice.

Finally, *how does advertising effectiveness change with time* and how can we model it?

Interactions. Is it better to advertise where sales are strong or weak? This is a classical argument, certain to draw proponents to each side. One can be sure that every model contains one or more, often inconspicuous, assumptions relating to this question, and so does any statistical analysis. In a similar vein, *are advertising effects additive with other marketing variables*, e.g., price, promotion, and competitive actions, *or multiplicative, or do they interact in more complicated ways*? All shades of assumptions appear in the model building and statistical literature. They are certainly not all consistent with one another.

2. BASIC PHENOMENA: WHAT CAN BE LEARNED FROM THE DATA?

Measurements must eventually resolve the issues just raised and tell

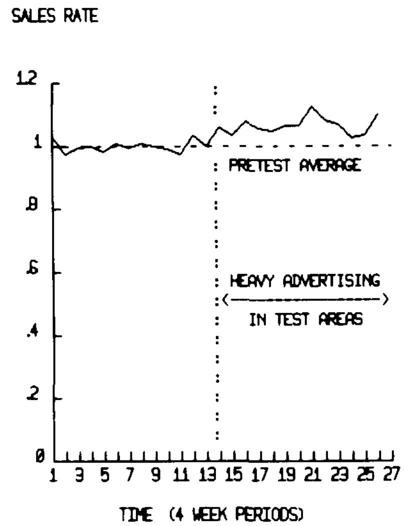


Figure 1. Sales rate of a packaged good responded quickly to increased advertising. Vertical axis shows the ratio of sales in test areas to sales in control areas, normalized to pretest average.

us which advertising phenomena are real and which are only folklore. In this spirit, we present a collection of empirical examples of certain major effects. These will help sort out the models in the next section.

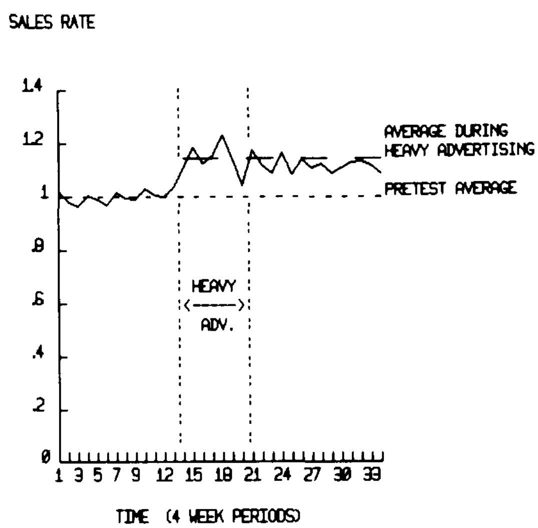


Figure 2. Sales rate of a packaged good rose quickly under increased advertising but declined slowly after it was removed. Vertical axis shows the ratio of sales in test areas to sales in control areas not receiving the heavy advertising.

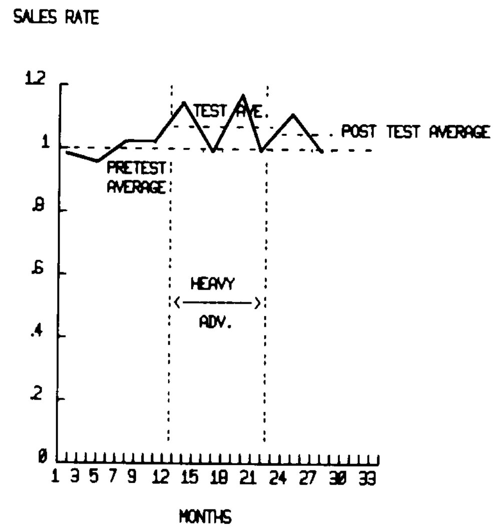


Figure 3. Another example of quick response upward followed by slow decay. Sales show more variance than Figure 2, but same general effect. Vertical axis is again normalized test/control.

Upward Response

Advertising increases sales, or such is the intent. Figures 1–3 show instances of sales before and after the introduction of substantial new advertising dollars. In each case the sales rate increases within a month

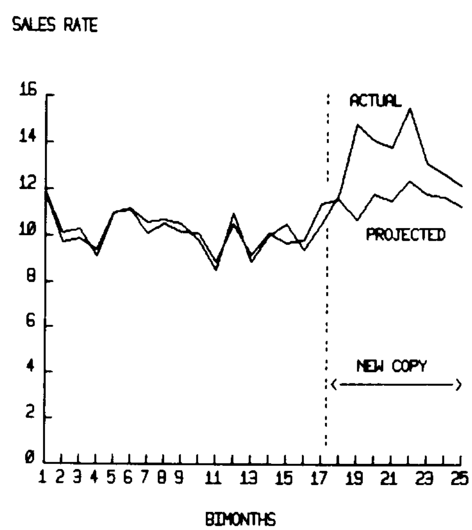


Figure 4. New advertising copy substantially increased Horlick's sales. Actual sales are compared to projections of a model based on market before copy change. Data replotted from Bloom, Jay and Twyman [9].

or two. Observe that this time span is, in each case, shorter than the judgments reported in the previous sections.

Figure 4, taken from data of Bloom, Jay and Twyman [9], is particularly interesting because it shows a jump in sales due, not to an increase in spending, but to a change in copy. Thus "advertising rate" is not necessarily the same as "spending rate." Notice again that sales respond almost immediately. A similar copy change effect appears in the results of Pekelman and Tse [48].

Sales at the New Level

Figures 1–3 show sales leveling off under the new, higher spending rates. Whatever was going to happen in these cases appears to have happened before the advertising stopped. Haley [20], however, has found a further effect shown in Figure 5. The sales increase occurs but its magnitude decreases with time. The leveling off appears to take place at a value lower than the initial gain. Such an effect is common in the case of new products that are purchased frequently. In such cases people learn about the product through advertising and try it, thereby causing a sharp spurt in sales. Only a fraction of the triers become regular purchasers and so sales taper off to a lower rate. In this paper we deal with established brands, but an analogous process seems quite likely: Increased advertising leads a group of nonusers to buy the product for reexamination or just for variety. Some of these customers continue to purchase, others not.

The copy induced sales increase in Figure 4 also seems to fall off. This too may be a new-trier effect, although many advertising people would say that the copy is wearing out. Such a description, however, seems more of a definition than an explanation.

Downward Response

Figures 2, 3, and 5*C* show sales response to decreased advertising. Notice that sales decay appears to take place more slowly than sales growth. This is particularly evident in Figure 2 and with more variance in Figure 3. In these cases we are able to observe the same product under both increases and decreases of advertising.

A posssible explanation for decay time being longer than rise time is that the rise relates to the advertising communications process, i.e., hearing or seeing the advertising message, absorbing it and acting on it. Since nominal forgetting times for advertising are on the order of a month (Lodish [36] and Strong [66]), it seems reasonable that an established product with good retail availability would show the positive effects of increased advertising within a short time. Krugman [26] argues that 3 exposures may be enough to stimulate action. On the other hand, decay in the absence of advertising seems more a question of experience with the product. Using and liking a brand will have far more influence on a customer than advertising. Although sales decay will depend on compet-

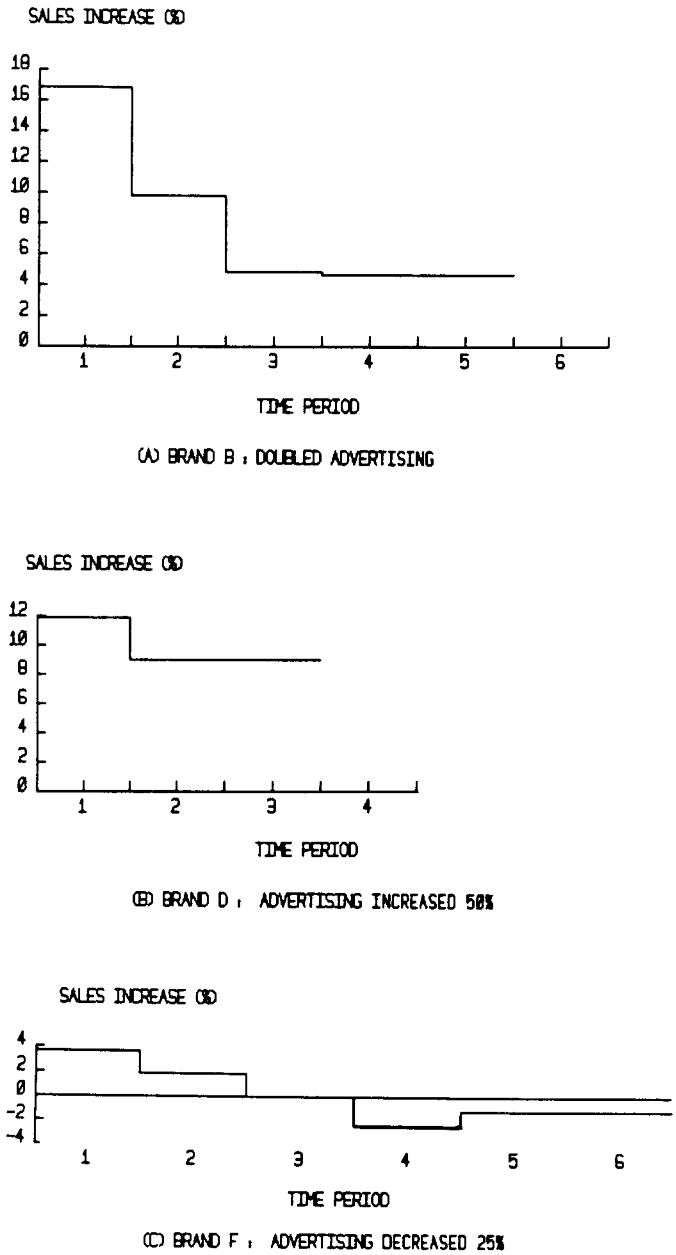


Figure 5. Field experiments reported by Haley [20] show the erosion of sales increases attained by heavier advertising (Brands B and D) and a relatively slow sales decay following decreased advertising (Brand F). Vertical axis displays the percent sales increase in test areas relative to control areas.

itive activity and other factors, it does not seem surprising (especially when facts stare us in the face) that decay is often much slower than growth.

An essential point, however, is that a good model of sales response to advertising should permit different rise and decay rates.

Sales with Zero Advertising

Figure 6 shows the sales of a line of never-advertised products. Many people do not realize this, but there are literally hundreds of unadvertised products selling happily away in every supermarket and department store. This often occurs when distribution is assured. Thus, chain-store house-brands are guaranteed a place on the shelf. Stores also stock unadvertised "price brands" with unfamiliar names in order to offer the consumer a low cost choice. In other examples, vending machines look out on a captive market and frequently carry unadvertised and virtually

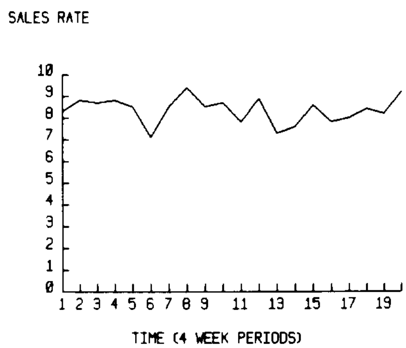


Figure 6. The healthy sales of a line of unadvertised food products show that advertising is not always required in order to sell something.

unknown brands. Department stores stock certain items by function without fanfare, e.g., string, envelopes or thumbtacks. This does not mean that such products would not sell faster with advertising but rather that positive sales with zero advertising are quite reasonable.

We should not be surprised, therefore, that empirical studies of sales response often indicate that a substantial part of the market seems not to be affected by advertising, at least over the medium run. This is noticeable in econometric studies with linear models where positive constant terms are common (e.g., Bass and Clarke [3]).

Thus an advertising response model should admit the possibility of sales with zero advertising (many do not).

Nonlinearity

Suppose advertising is held constant and other market conditions do not change. After some time period the market can be expected to be in steady state. If this were done for a number of different advertising rates, we could make a plot of steady state sales vs. advertising.

What would the curve look like? We would not expect it to be linear, for this would have a variety of nonsensical consequences. (For example, a product with a fixed production cost per unit would have an optimal advertising rate of either zero or infinity.) However, "not linear" covers many possibilities. We describe two important ones.

Diminishing Returns. Figure 7 displays a pair of empirical advertising response curves plotted from data of Benjamin and Maitland [8]. Their data are particularly valuable because of the great range of advertising levels studied. In each case the slope of the curve decreases at high advertising levels, thereby showing concavity or diminishing returns. Less obvious is whether response is better modeled by an absolute ceiling

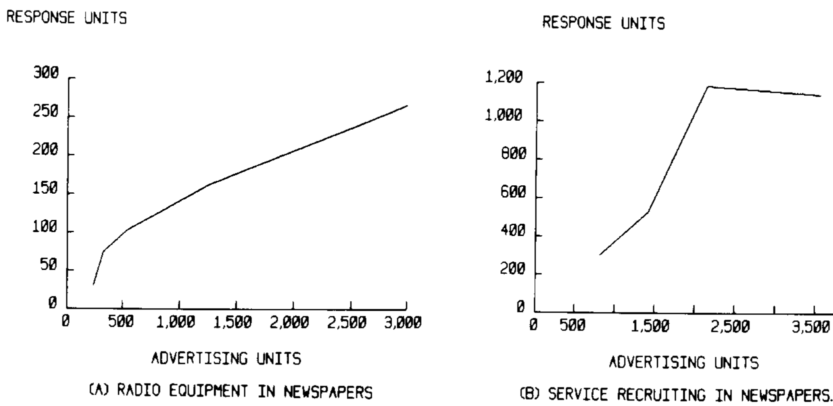


Figure 7. Two examples of nonlinear response exhibit the phenomenon of diminishing returns at high advertising rates. Data replotted from Benjamin and Maitland [8].

(saturation level) or by a function that can surpass any prespecified level, albeit with increasing difficulty. Benjamin and Maitland [8] choose the latter course; they take sales to be the log of advertising. Such a function, however, does not make sense at zero advertising since $\log 0 = -\infty$.

S-Shape. Controversy surrounds the question of whether steady-state sales response to advertising is S-shaped, i.e., whether, at low levels of advertising, increases are increasingly effective up to some point after which diminishing returns set in. Simon [64], for example, says no.

However, as mentioned earlier, many advertising schedules today contain *flights* or *pulses.* A theory that might justify flights is that response is S-shaped, e.g., small advertising rates do little good but medium rates are effective. Published empirical evidence of such relationships is hard to find.

Rao [51] and Rao and Miller [53] display S-shaped response curves which are developed in a three-step process. First Rao analyzes individual market areas by times-series regressions, relating sales to advertising with a linear model. The coefficient of advertising is the slope of the sales vs. advertising curve of the linear model. Often, but not always, he finds a small slope (i.e., less advertising effectiveness) where the average advertising in the market is either very low or very high (Figure 8*A*). As a second step Rao fits a curve cross-sectionally through the slope data to obtain a relationship between slope and average advertising rate. This is often quadratic. The final step is to integrate the slope vs. advertising relation to obtain sales vs. advertising, i.e., the sales response curve. If the slope is quadratic as in Figure 8*A*, the sales response will be an S-shaped cubic over the relevant range as shown in Figure 8*B*. Rao [51] has three or four examples that show generally similar results but also reports that some products show no S-shape.

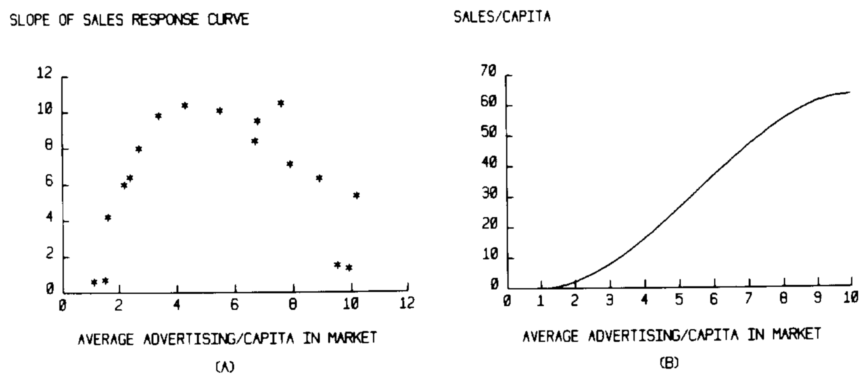


Figure 8. An S-shaped sales response curve developed by Rao's [53] method for a canned juice [51]. (*A*) Individual market areas analyzed by time series show different advertising response coefficients (slopes). Each point is one market area. (*B*) Integration of a quadratic fit to (*A*) yields S-shaped sales response curve. Scaling is arbitrary.

In a similar vein Wittink [74] does time series analyses of individual markets and then cross-sectional studies of the advertising slope coefficients. He too finds larger slopes at larger advertising rates, indicating the lower region of an S-curve. The upper part is presumably guaranteed by ultimate saturation of advertising effect.

On the direct question of the efficacy of pulses (as opposed to whether steady-state response is S-shaped), Ackoff and Emshoff [1] report good results from pulsing. Sethi [58] reports a Milwaukee Advertising Laboratory experiment that seems to show good short run but poor long run effects. In any case, considering current practice, Rao and Miller's work, and the importance of the issue, we argue that advertising models should accomodate S-shaped curves.

Before leaving the empirical evidence on shape, we present certain provocative results from McDonald [38]. He has analyzed panel data that contained not only product purchases but also media exposure. Figure 9 shows a sales measure plotted against an advertising measure. The sales measure is the percentage of brand switches *to* the advertised product as a proportion of switches both *to* and *from* it. Thus 50% would be expected in the absence of an advertising effect and, in fact, Figure 9 averages to 50% if each point is weighted by its number of observations. The advertising measure is the number of opportunities to see ads for the brand in the last four days of the customer's time interval between successive purchases. The curve is an aggregate over several product classes and many brands, all essentially supermarket items. The curve is not comparable to those presented earlier because it deals with individuals, not total market, and because both time interval and sales measure are very

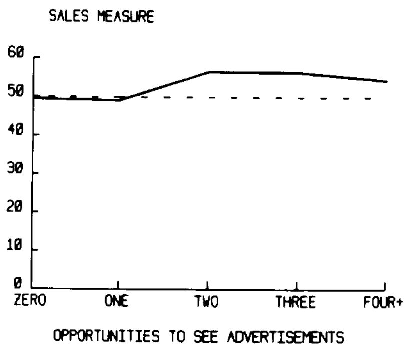


Figure 9. Microlevel evidence for S-shaped sales response is obtained from observing individuals. The sales measure is the number of switches to the advertised brand as a percentage of the total number of switches to and from the brand. Opportunities-to-see include only those in the last 4 days prior to purchase (McDonald [38]).

specialized. However, the results are quite revealing, especially the S-shape, the seeming saturation after just a few exposures and the evidence of immediate advertising effects.

Impulse Response

A standard question about a dynamic system is, "What is its impulse response?" Thus, suppose we put a short burst of advertising into the market, say an expensive TV special, a multipage four-color spread in a magazine, or a massive direct mail drop; what would be the resulting shape of the sales response over time?

Figure 10 shows an example of this. A test group of people was exposed and a control group not exposed to a sharp pulse of advertising. The ratio of test sales/control sales in the following months was recorded. A number of tests have been averaged to give the impulse response shown.

Another type of analysis, common in econometric studies, measures the effect of past advertising on current sales by regression. This yields

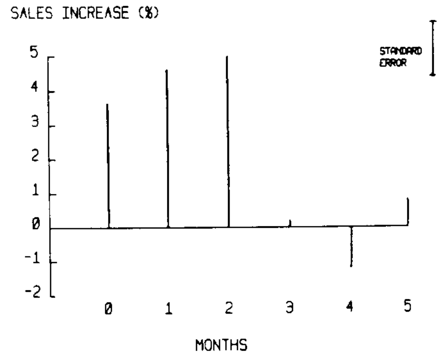


Figure 10. A large impulse of advertising in month zero yields substantial sales increases in months 0, 1 and 2 for an infrequently purchased consumer durable.

an implied impulse response even though the advertising was not actually done in pulses. Figure 11, plotted from the results of Bass and Clarke [3], displays such a case.

Notice that Figures 9–11 corroborate earlier observations that response to advertising is relatively quick. The initial effect of a pulse takes place within 2 months. This is in line with the rise times in Figures 1–3. Ideally, impulse response measurements would also pick up long run effects in the tail. However, if the decay is as slow as those of Figures 2 and 3, the usual statistical methods will have difficulty detecting it.

In examining alternative models in the next section we can determine their impulse responses and compare them to what we observe empirically.

Infrequent Purchases

Figure 10 is especially interesting because it deals with a consumer

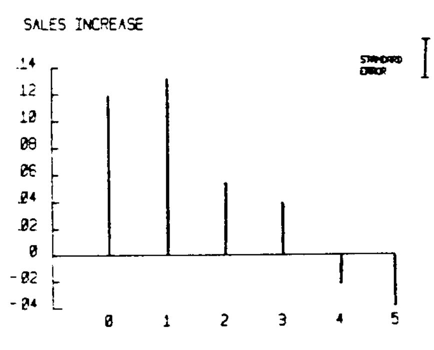


Figure 11. A unit impulse of advertising in month zero produces sales in that and subsequent months. Results determined by econometric methods for a dietary weight control product by Bass and Clarke [3].

durable whose normal time between purchases is measured in years. Some people have argued that the fast advertising response discussed earlier will not apply to infrequently purchased goods. Figure 10 refutes this. The reason such goods can respond quickly is simple enough. At any given point of time some people are in the market, ready to act. Indeed, potential customers often seek information and take a special interest in the advertising for the product class.

However, for infrequent or one-time purchases like houses, refrigerators, books, college educations, or enlistments in the armed services, a new phenomenon is likely to come in: *market depletion*. Figure 12, taken from data of Benjamin, Jolly and Maitland [7], displays the effect. Successive advertisements in a periodical draw fewer and fewer customers, tending toward an asymptotic value where market depletion is

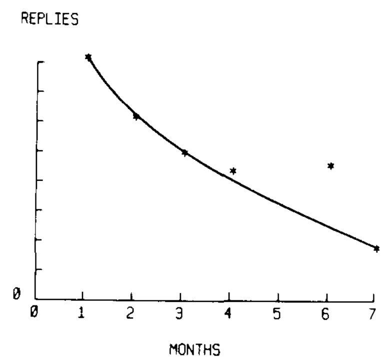


Figure 12. Replies to a series of advertisements in a periodical show evidence of market depletion and temporary rejuvenation. Replotted from Benjamin, Jolly and Maitland [7].

balanced by new entry, or zero if there is none. The authors also observed that when an advertisement was omitted the next one met increased response, indicating a degree of market rejuvenation.

Although statistical significance is not there, the impulse response curves of Figures 10 and 11 hint at a negative sales reaction about four months after the advertising pulse. Such borrowing of future sales is a type of temporary market depletion often found in consumer promotions and one that undoubtedly sometimes occurs with advertising. Becknell and McIsaac [5], for example, report the effect in cookware.

Competition

Companies worry about competition. Surely, if one brand can increase its sales and share by advertising, so can another. Therefore, one brand's advertising will often reduce another brand's sales. Some researchers have studied this phenomenon, for example, Lambin, Naert and Bultez

[31] and Horsky [23]. Figure 13 shows curves derived from data of the former. We argue that an understanding of advertising phenomenon in consumer markets requires competitive models.

Issues Outstanding

For a number of questions raised earlier, straightforward evidence is scanty.

Advertise Where Sales are Strong or Weak? Undoubtedly this question is too simplistic and the right answer depends on conditions. One might expect, for instance, that advertising response would be poor where distribution is weak. On the other hand a concerted marketing program

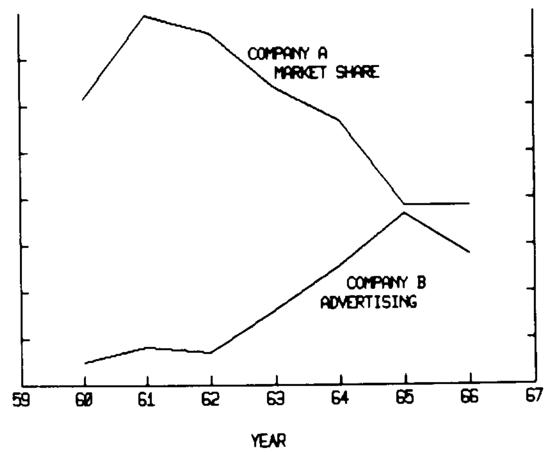


Figure 13. Company A loses share to aggressive Company B marketing that includes heavy advertising. The product is a low price consumer durable. Data derived from Lambin, Naert and Bultez [31].

that includes substantial advertising may be required to gain distribution and the benefits beyond.

The influencing conditions are likely to vary from case to case. Haley [20] produces evidence for better response where sales are already increasing. Rao and Miller [53] report a product for which advertising response is greater where share is greater but Wittink [74] reports the reverse. Competitive advertising can affect response. The various conditions need sorting out.

Hysteresis. Are there situations for established products where advertising can carry sales up to new levels to stay there after advertising is reduced? Parsons [45] explores what appears to be such a case, but good examples are not generally available.

Interactions. How does advertising interact with other marketing variables? Some models assume additive effects, some multiplicative, and others more complicated relationships. They cannot all be right in a given situation. Interactions are usually much harder to measure than main effects, although some start has been made (Swinyard and Ray [67], Eskin and Baron [14] and Wildt [73]). Almost certainly, advertising response varies over the product life cycle. Some studies have found that advertising response for a product differs from market area to market area. This may result from different product class strength, demographic segmentation, or distribution levels. Much unraveling needs to be done.

Conclusions

The empirical evidence suggests that at least the following phenomena should be considered in building dynamic models of advertising response:

P1. Sales respond dynamically upward and downward to increases and decreases of advertising and frequently do so at different rates.

P2. Steady-state response can be concave or S-shaped and will often have positive sales at zero advertising.

P3. Competitive advertising affects sales.

P4. The dollar effectiveness of advertising can change over time as the result of changes in media, copy, and other factors.

P5. Products sometimes respond to increased advertising with a sales increase that falls off even as advertising is held constant.

All of these effects hold implications for managerial action. Obviously other important phenomena also exist, some of which have been discussed and others of which remain to be discovered. However, parsimony prompts us to keep the list short.

We now look for models that embrace these basic elements. The list does not seem very demanding, and indeed, where there are competing ways to represent the same phenomenon, we shall not be well equipped to distinguish among them. However, even our simple requirements of face validity will find many models wanting.

3. MODELS

For twenty years researchers have been adding marketing models to the literature like grains of sand to the beach. By now the pile, if not a dune, is at least a sand castle. Two rather dramatically different model building traditions coexist uneasily in the literature. One, which we shall call *a priori*, draws heavily on intuition and, although its practitioners are not oblivious to data, the model building goal is to postulate a general structure, not describe a specific application. In this category we place Vidale and Wolfe [71], Nerlove and Arrow [43] and Little [33, 34]. The other tradition is statistical or *econometric* and usually starts from a specific data base, e.g., time series of sales or share and advertising. In this category are Bass [2], Bass and Clarke [3], Montgomery and Silk [41] and Lambin [30] to name a few. In addition some older work and an increasing amount of new work is *mixed* in that it starts with rather more complicated a priori models and endeavors by statistical methods to fit and evaluate them. Examples are Kuehn, McGuire and Weiss [28] and Horsky [23].

A Priori Models

Vidale-Wolfe. In 1957 Vidale and Wolfe [71] published one of the earliest and most interesting of advertising response models. They used three basic ideas: (1) sales rate increases with advertising rate, (2) this effect decreases as sales rate approached a value called saturation and (3) sales constantly erode spontaneously. The authors give empirical illustrations of these phenomena. Let s = sales rate (sales units/period), $\dot{s} = ds/dt$, x = advertising rate (dol/period), ρ = response constant (sales units/dol/period), λ = decay constant (period^{-1}) and m = saturation sales rate (sales units/period). Sales might be measured in kilograms, liters, pounds, cases, etc., periods in weeks, months, years, etc.

The Vidale-Wolfe structure is

$$\dot{s} = \rho x[1 - (s/m)] - \lambda s. \tag{1}$$

The model contains only three constants, yet displays many of the characteristics one would intuitively attribute to advertising response. Since (1) is a first order ordinary differential equation, it has an explicit solution for arbitrary $x(t)$. We shall report it for completeness, but for more intuitive understanding, we shall display (a) sales response to a rectangular pulse, (b) impulse response and (c) steady state response. Suppose that at $t = 0$, $s = s(0)$, and a constant rate of advertising $x(t) = x$ is started which lasts until $t = T$ when it drops to zero. Solving (1) for such a rectangular pulse yields

$$s(t) = \begin{cases} r(x) + [s(0) - r(x)]e^{-[1+(\rho x/\lambda m)]\lambda t} & 0 \leq t \leq T \\ s(T)e^{-\lambda(t-T)} & T < t \end{cases} \tag{2}$$

where

$$r(x) = m(\rho x/\lambda m)/[1 + (\rho x/\lambda m)]. \tag{3}$$

Equation (2) is sketched in Figure 14*a*. Notice that the rise time is primarily affected by the constant ρ and decay time by λ.

The impulse response, expressed as the incremental sales generated by an amount, X, of dollars spent in a very short time at $t = 0$, is

$$\begin{aligned} \Delta s(t) &= s(t) - s(0)e^{-\lambda t} \\ &= [m - s(0)][1 - e^{-\rho X/m}]e^{-\lambda t}, \qquad 0 < t \end{aligned} \tag{4}$$

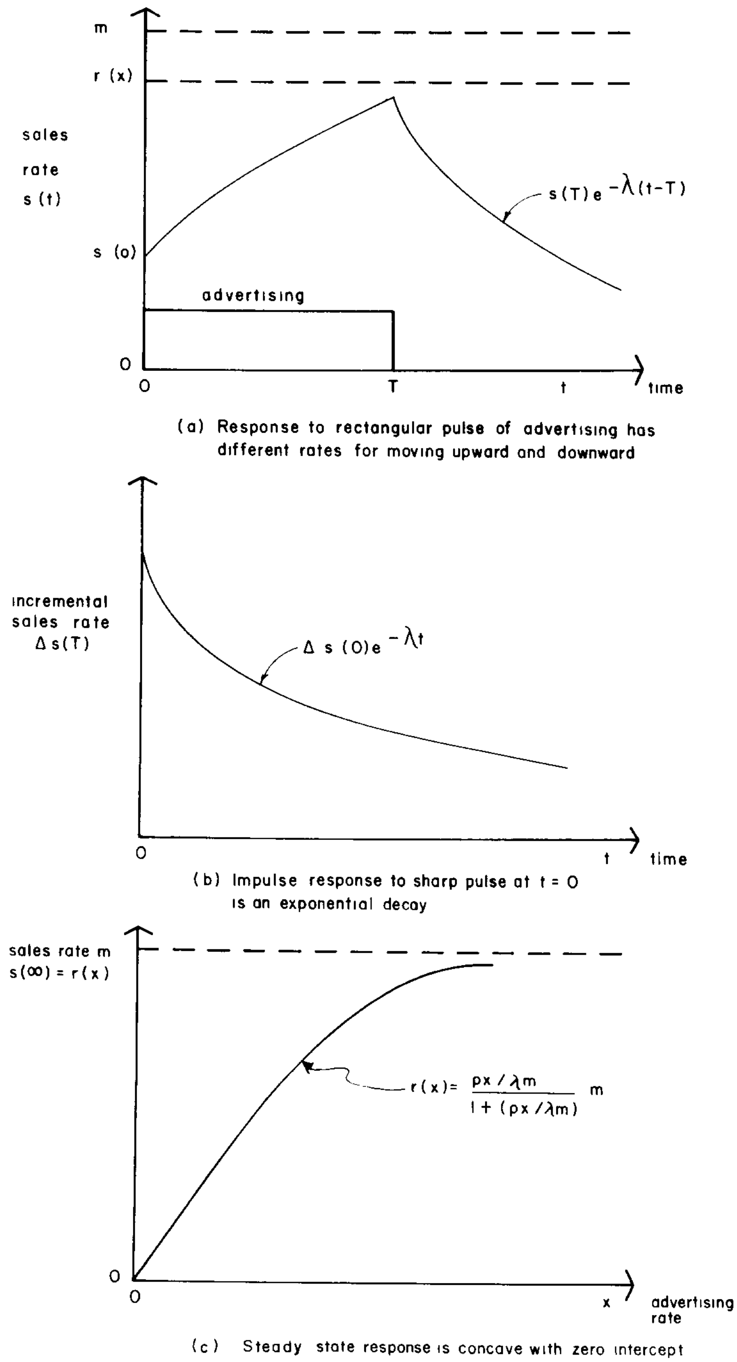


Figure 14. Vidale-Wolfe model: Sales response to advertising.

and is sketched in Figure 14*b*. Impulse response is exponential with decay constant λ.

The steady state response to a constant advertising rate x, is

$$s(\infty) = r(x) \tag{5}$$

with $r(x)$ given by (3) and sketched in Figure 14*c*.

The general solution to the Vidale-Wolfe differential equation for arbitrary $x(t)$ is:

$$s(t) = \left\{\int_0^t \left[\exp \lambda \int_0^u (1 + \rho x(v)/m\lambda)dv\right]\rho x(u)du + s(0)\right\} \cdot \exp\left\{-\lambda \int_0^t (1 + \rho x(u)/m\lambda)du\right\}.$$

In comparing the Vidale-Wolfe model with our catalog of phenomena, we find that it has different rise and decay times in good agreement with P1. Steady state response, however, is concave, cannot be S-shaped and has zero sales at zero advertising. This is not the flexibility called for by P2. The model does not consider competitive advertising in disagreement with P3. No explicit provision is made for changes in copy or media effectiveness as required by P4, although ρ could be made to perform some of that role. The temporary sales increases of P5 are not handled. The exponential impulse response corresponds weakly to Figures 10 and 11.

Nerlove/Arrow. In a 1962 study of advertising dynamics Nerlove and Arrow [43] employ the term "goodwill," which "summarizes the effects of current and past advertising outlays on demand." Let A = stock of goodwill (dollars), x = advertising rate (dol/period), $\dot{A} = dA/dt$ (dol/period), δ = goodwill depreciation rate (1/period). They postulate that growth and decay of goodwill behave according to

$$\dot{A} = x - \delta A \tag{6}$$

Goodwill, price and other variables affect sales. Let p = price (dol/unit), z = variables uncontrolled by the firm, $s = s(p, A, z)$ = sales rate (units/period). The authors' purpose is to investigate mathematical conditions required of optimal policies under various circumstances.

Our interest is in sales response. Since sales is presumably a monotone transformation of goodwill, the shape of rectangular, impulse and steady state response for sales will closely depend on that for goodwill. Response to a rectangular advertising input, $x(t) = x$ for $0 \le t \le T$ and $x(t) = 0$ for $t > T$ is

$$A(t) = \begin{cases} A(0)e^{-\delta t} + (x/\delta)[1 - e^{-\delta t}] & 0 \le t \le T \\ A(T)e^{-\delta(t-T)} & T \le t \end{cases} \tag{7}$$

Incremental response to an impulse of X dollars administered at $t = 0$ is exponential:

$$\Delta A(t) = A(t) - A(0)e^{-\delta t} = Xe^{-\delta t} \qquad 0 < t \tag{8}$$

Steady state response to constant $x(t) = x$ is linear:

$$A(\infty) = x/\delta \tag{9}$$

At a later stage of their paper, Nerlove and Arrow investigate the constant elasticity response function, $s = kp^{-\eta}A^{\beta}z^{\xi}$, which, for present purposes, can be written

$$s(t) = kA(t)^{\beta} \tag{10}$$

with $\beta < 1$ for meaningful functions. Figure 15 sketches rectangular, impulse, and steady state sales responses.

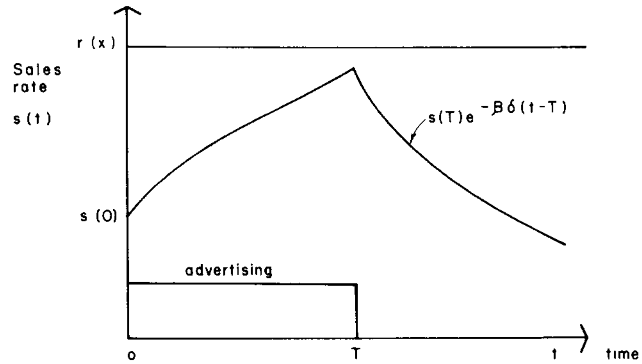
The Nerlove-Arrow model views advertising as piling up goodwill, which continuously leaks away. The current stock of goodwill drives a steady state response function, exemplified as a constant elasticity model. The process is somewhat similar to the Vidale-Wolfe model but the latter differentiates between rise and decay, whereas Nerlove-Arrow does not, since one constant, δ, governs both processes. Thus the model fails on the key phenomenon P1. The constant elasticity model for steady state response has the problem of zero sales at zero advertising and lacks the possibility of an S-shape, thereby lacking the flexibility of P2. There is no consideration of competition (P3), changing effectiveness (P4), or temporary sales increases (P5). With appropriate changes of functions and parameters, most likely P2–P4 could be accommodated but not P1 or P5. The authors give no empirical evidence for their model.

Lanchester Models. We shall give the name Lanchester to a flexible class of competitive marketing models that have a strong resemblance to Lanchester's models of warfare. The basic idea was introduced in 1957 by Kimball [25]. A model of this form has also been considered by Deal and Zionts [13] and a closely related discrete-time version by Schmalensee [57]. We concentrate on a basic two-competitor case and later point out certain generalizations. Let s_1 = sales rate of brand 1 (units/period), s_2 = sales rate of brand 2 (units/period), x_1 = advertising rate of brand 1 (dol/period), x_2 = advertising rate of brand 2 (dol/period), ρ_1 = advertising effectiveness constant of brand 1 (dol^{-1}), ρ_2 = advertising effectiveness constant of brand 2 (dol^{-1}) and m = total market sales rate (units/period):

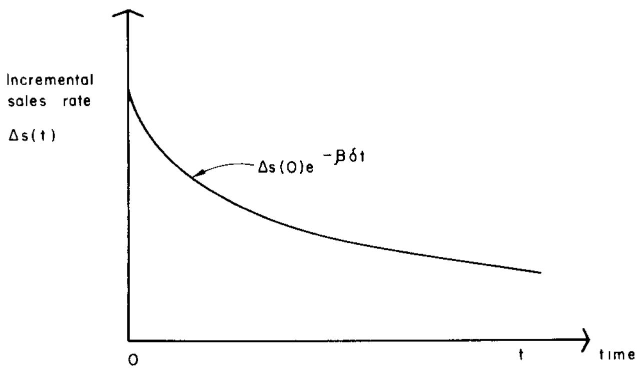
$$s_1 + s_2 = m. \tag{11}$$

The basic Lanchester model is

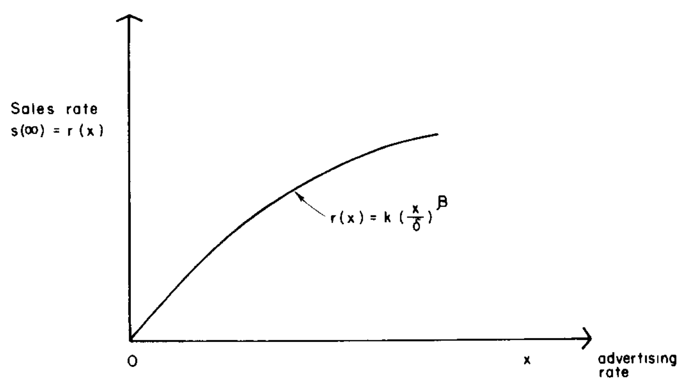
$$\begin{aligned} \dot{s}_1 &= \rho_1 x_1 s_2 - \rho_2 x_2 s_1 \\ \dot{s}_2 &= \rho_2 x_2 s_1 - \rho_1 x_1 s_2. \end{aligned} \tag{12}$$



(a) Response to rectangular pulse of advertising has upward and downward movement linked through δ



(b) Impulse response to sharp pulse at $t=0$ is an exponential decay



(c) Steady state response is concave ($\beta<1$) with zero intercept

Figure 15. Nerlove-Arrow model: Sales response to advertising with constant elasticity response function.

Thus, Company 1 wins sales proportional to its advertising and to Company 2's sales. At the same time Company 1 is losing sales proportional to its own sales and Company 2's advertising. The situation is entirely symmetric for Company 2. The coefficients ρ_1 and ρ_2 permit different advertising dollar efficiencies due to copy, media buying, and other product and market characteristics.

A number of interesting properties of the model emerge from simple analyses. First, we make the substitutions:

$$s_2 = m - s_1 \tag{13a}$$

$$\rho = \rho_1 m \tag{13b}$$

$$\lambda = \rho_2 x_2. \tag{13c}$$

Dropping the now redundant subscript 1, we obtain $\dot{s} = \rho x[1 - (s/m)] - \lambda s$, which is just the Vidale-Wolfe model. Thus, the Lanchester equations (12) form a competitive generalization of Vidale-Wolfe. Note that the decay constant of the Vidale-Wolfe model is now expressed in terms of the competitor's advertising rate.

It follows that, for the case of fixed competitive advertising, appropriate substitutions into (2) to (5) give the rectangular pulse, impulse, and steady state responses and Figure 14 portrays their shapes. The case of time-varying advertising and/or time-varying competitive advertising converts into a first order differential equation which can be solved explicitly if desired.

The steady state response functions help build intuition about the competitive effects of advertising. Solving (11) and (12) yields

$$\begin{aligned} s_1 &= m(\rho_1 x_1)/(\rho_1 x_1 + \rho_2 x_2) \\ s_2 &= m(\rho_2 x_2)/(\rho_1 x_1 + \rho_2 x_2). \end{aligned} \tag{14}$$

Of great interest is the property that one company's response function depends on another company's advertising rate. This is sketched in Figure 16.

Response models of the general type us/(us + them) are well known. In particular Friedman [15] Mills [40] and Bell, Keeney and Little [6] study them. These papers refer to generalizations to N competitors, other functions of advertising, various game theoretic issues, and generalizations beyond advertising. A straightforward expansion of (12) to N competitors with x_j generalized to $x_j^{\epsilon_j}$ produces a model with many of the requested phenomena:

$$\dot{s}_i = \rho_i x_i^{\epsilon_i} \textstyle\sum_{j \neq i} s_j - (\sum_{j \neq i} \rho_j x_j^{\epsilon_j}) s_i \qquad i = 1, \cdots, N \tag{15}$$

$$\textstyle\sum_{j=1}^{N} s_i = m. \tag{16}$$

In steady state

$$s_i = m\rho_i x_i^{\epsilon_i} / \sum_{j=1}^{N} \rho_j x_j^{\epsilon_j} \qquad i = 1, \cdots, N \tag{17}$$

The response function (17) is quite versatile, being S-shaped in x_i for $\epsilon_i > 1$ and concave for $0 \le \epsilon_i \le 1$. Thus, if we think of the ρ_i as carrying media and copy effectiveness, the Lanchester model (15–17) displays phenomena P1–P4 except for nonzero sales at zero advertising. The latter might be treated, at least in principle, by an additive constant. Or, expressed another way, we could say in this and other models that we are dealing with the "advertising affectable market." The model does not display P5, erosion of incremental sales under constant advertising.

A further generalization would be to make each brand's advertising differentially effective against each other brand, e.g., change $\rho_i x_i^{\epsilon_i} s_j$ to

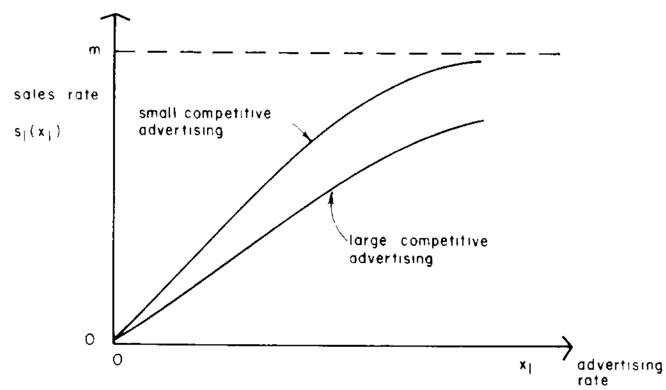


Figure 16. Steady state sales response is affected by competitive advertising in a Lanchester model.

$\rho_{ij} x_i^{\epsilon_i} s_j$. Another feature would be to let the total market size m depend on total industry advertising.

Brandaid. Little [34] presents a general, flexible structure for modeling the effect of the marketing-mix on company sales. The advertising submodel works as follows. Let t = time in discrete units (periods), $s(t)$ = brand sales rate (units/period), $a(t)$ = brand advertising rate (index), $r(a)$ = long run (steady-state) advertising response (units/period) and $\alpha(a)$ = carry-over constant. Customer purchases are presumed to have persistence so that current sales are a weighted combination of previous sales and long run response.

$$s(t) = \alpha s(t-1) + (1-\alpha) r(a(t)). \tag{18}$$

Steady state response is arbitrary; in particular, it can be S-shaped and have a nonzero origin as sketched in Figure 17. The burden of calibration is placed on the user. In applications to date some companies have made empirical measurements that guide the setting of $r(a)$ and some have used managerial judgment or a mix of the two.

The model anticipates that media and copy effectiveness may vary over time. Advertising consists of *messages* delivered to individuals by *exposures* in *media* paid for by *dollars*. These ideas are modeled by

advertising rate
= (copy effectiveness) × (media efficiency) × (spending rate).

Let $h(t)$ be copy effectiveness, $k(t)$ media efficiency, $x(t)$ spending rate, and h_0, k_0 and x_0 normalizing constants for these quantities. Then the advertising rate, $a(t)$, is given by

$$a(t) = h(t)k(t)x(t)/h_0 k_0 x_0. \tag{19}$$

This quantity can drive the response function or, as a further embellish-

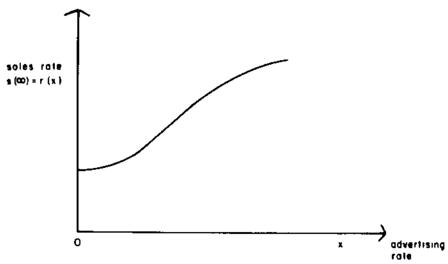


Figure 17. The Brandaid advertising submodel accommodates arbitrary steady state response curves, e.g., an S-shaped curve with nonzero intercept.

ment, a weighted combination of current and past advertising can be used. A simple exponential smoothing model is

$$\hat{a}(t) = \beta\hat{a}(t-1) + (1-\beta)a(t) \tag{20}$$

where $\hat{a}(t)$ is the effective advertising at t and β is a carryover constant for advertising exposure (units of fraction/period, $0 \leq \beta \leq 1$).

Brandaid treats copy changes by letting a coefficient that multiplies advertising spending vary with time. The same approach can be applied to most of the other models discussed in this paper. The technique seems appropriate for copy changes that communicate the same information better, for example, by using a more attention-getting format or providing a better demonstration of product attributes. Such changes make the advertising dollar go further and so can be considered to affect the scaling of the advertising axis of the sales response curve. On the other hand, copy changes that reposition the product and appeal to different groups of people or suggest new uses affect the sales axis. The basic sales potential of the product may go up or down. This is not modeled explicitly in Brandaid or any other model discussed here, but such a step might be a good one.

We see, therefore, that the Brandaid advertising model meets the criteria of flexibility in dynamic and steady state response (P1, P2) and treats changing effectiveness (P4) to the extent of a changing coefficient. The Brandaid paper also presents a way to model competition that lends itself well to calibration by managerial judgment in decision calculus style but seems less suited to our purposes here. The model has no mechanism for handling the temporary sales increase phenomenon P5.

We now search for unification and show that the previous three models are special cases of Brandaid; or, if you prefer, the previous models have gone out on a limb with specific postulates where Brandaid has refused to make commitments.

Consider first the Vidale-Wolfe model. We convert it to discrete time by the approximation

$$\dot{s} \cong [s(t) - s(t-h)]/h \tag{21}$$

where h is a small interval of time. Taking the time unit equal to h (i.e. setting $h = 1$) and defining

$$\alpha(x) = 1/[1 + \lambda + (\rho x/m)] \tag{22a}$$

$$r(x) = (\rho x/\lambda m)/[1 + (\rho x/\lambda m)] \tag{22b}$$

we obtain, by substituting (21) and (22) into (1) and rearranging, $s(t) = \alpha(x)s(t-1) + [1 - \alpha(x)]r(x)$. This is just the Brandaid advertising model with $a(t) = x(t)$, the spending rate. Notice in (22) that $0 < \alpha < 1$ and that $r(x)$ is indeed the steady state response of the Vidale-Wolfe model.

The implications of the relation between the two models are several. First, by appropriate specification of $\alpha(x)$, Brandaid can have different rise and decay times. Second, the Brandaid advertising model turns into "our brand" of a two-brand discrete time Lanchester model through substitution of (13b) and (13c) into (22). N competitor generalizations are also possible so that in fact the Lanchester model (15) can be cast into the same form.

The Nerlove-Arrow model in discrete time is a special case too. Set $\alpha = 0$ in (18), suppress $h(t)$ and $k(t)$ in (19) and drive the response function in (18) with the effective advertising $\hat{a}(t)$ of (20). Effective advertising corresponds to Nerlove and Arrow's goodwill.

Finally, we note two straightforward generalizations of the lag structure. For (18) $s(t) = \sum_{i=1}^{\infty} \alpha_i s(t-i) + (1 - \sum_{i=1}^{\infty} \alpha_i) r(a(t))$, and for (20) $\hat{a}(t) = \sum_{i=0}^{\infty} \beta_i a(t-i)$ where $\sum_{i=0}^{\infty} \beta_i = 1$. These generalizations are not especially parsimonious as each added parameter puts more burden on calibration. A situation in which additional sales lags might be desired is when sales are measured by factory shipments so that the distribution pipelines put delays between customer purchase and point of measurement.

Other Models. The literature contains a variety of other a priori models, a number of which we report here.

Sasieni [54] postulates sales dynamics in the form $\dot{s} = g(s, x, t)$ where g is a known function that increases with advertising, x, and decreases with sales, s ($\partial g/\partial x \geq 0$, $\partial g/\partial s \leq 0$). Vidale-Wolfe (1) is a special case. Schmalensee [56] goes a step further by postulating that, at every moment, there is an equilibrium demand toward which actual sales are moving. Equilibrium demand corresponds to our steady state sales rate with the addition that, in principle, the equilibrium point can change with time. In our notation, let $r = r(x, p, t)$ be the steady state sales rate as a function of advertising, x, price, p, and possibly t. Schmalensee investigates $\dot{s} = F[r(x, p, t), s(t)]$ and assumes $\partial F/\partial r > 0$ and $\partial F/\partial s < 0$. Again, Vidale-Wolfe can be cast in this form, using (1) and (3): $\dot{s} = [\lambda/(1 - r/m)](r - s)$.

The Brandaid advertising model fits into Sasieni's form but not quite into Schmalensee's. In continuous time Brandaid becomes

$$\dot{s} = \gamma(x)[r(x) - s] \tag{23}$$

where $\gamma(x) = \lim_{h\to 0}[1 - \alpha(x)]/h$ is the carryover function converted to a decay factor. The existence of $\gamma(x)$ keeps (23) from being in Schmalensee's form.

Sasieni and Schmalensee each have as a goal the characterization of optimal policies and so make as few assumptions as possible about response. This leads to very general formulations. Both are quite flexible on response upward and downward and on the shape of steady response. At the same time this means they specify relatively little about the mechanisms of advertising. Sasieni does not explicitly consider competition. Schmalensee introduces it only to the extent of formally indicating a competitive advertising variable in the equilibrium demand function.

A variety of generalizations and modifications of the Vidale-Wolfe and Nerlove-Arrow models have been proposed. Mann [37] generalizes the Nerlove-Arrow exponential weighting of past advertising for determining goodwill to more arbitrary weightings. Sethi, Turner and Newman [61] do approximately the same thing to Vidale-Wolfe. They introduce a variable termed market attitude determined by present and past advertising. Current advertising is thereby replaced in the model by a linearly weighted combination of present and past advertising.

Sethi [60] proposes a model $\dot{s} = \rho \log x - \lambda s$ which exchanges the

Vidale-Wolfe sales saturation process in (1) for a log function. Steady state response now becomes the strictly concave function $s = (\rho/\lambda)\log x$. From the point of view of our catalog of phenomena this has about the same advantages and disadvantages as Vidale-Wolfe except for the added drawback that the log model makes no sense at zero advertising. Burdet and Sethi [10] also present a discrete time model of Brandaid form with linear steady state response, an undesirable feature.

In the early and mid-1960s researchers created many speculative and often interesting models. Kuehn [27] presents a general marketing mix model motivated by the linear learning description of brand switching. Viewed as an advertising response model, sales consist of a retained fraction of past sales plus new input. The new input is linear in the brand's share of total advertising and in the brand's share of various interaction functions between advertising and other marketing variables. Shakun [62] gives a competitive model in which a firm's market share is share of total advertising but each firm's expenditure is weighted by its market share from the previous period. Industry sales of the product category are a saturating function of effective industry advertising. This in turn is a weighted combination of past effective advertising and new spending, diminished possibly by the spending on competing categories of products. Gupta and Krishnan [19] define effective advertising as a linear weighting of past advertising. Then, in a competitive model, market share equals company share of total effective advertising.

These models are all competitive and so satisfy our phenomenon P3. However, from the vantage point of today, they lack flexibility in rise and decay rate (P1) and have rather inflexible concave steady state response functions (P2).

In a totally different direction of development, Tapiero [68] studies a diffusion model of sales response to advertising. The model views sales as uncertain and the result of a stochastic process. However, the underlying response dynamics are basically Vidale-Wolfe. In still another approach Gould [18] describes the advertising process as a diffusion of information among individuals. His resulting differential equation is identical to Vidale-Wolfe.

Econometric Models

Whereas one group of researchers has proposed and promoted a priori models, another has embraced specific data bases and applied econometric methods to them. Parsons and Schultz [47] describe many of the techniques. The amount of econometric work is large. Clarke [11] finds more than 70 studies and, at that, restricts himself to those amenable to inferences about the cumulative effect of advertising. Lambin [30] alone analyzes 107 brands and reports 291 regressions.

Such studies take the historical data as it comes. The data may or may not contain sufficiently clean changes in advertising to draw solid inferences. Notice that most of our earlier examples of advertising phenomena were drawn from field experiments. It is easier to identify specific effects by direct manipulations than by sifting through the historical record with an econometric sieve. The drawback of experiments, of course, is that they require considerable effort to mount.

Most of the econometric studies use models that are linear or linear in the logarithms of the variables, with or without lagging some of them. Simultaneous equation models are common. Researchers add explanatory variables as available, e.g., other marketing activities, economic indicators, and dummy variables for special circumstances. We examine several major classes of econometric work, focusing, however, only on the advertising response models therein.

Linear in Advertising. Let s_t = sales in period t (sales units). x_t = advertising in period t (dollars). a_i, b_i = constants. A parsimonious linear model used, for example, by Bass and Clarke [3] is:

$$s_t = a_0 + \textstyle\sum_{i=0}^{L} b_i x_{t-i} \qquad (24)$$

The model has a linear steady state response, given by $s = a_0 + (\sum b_i)x$ and an arbitrary impulse response, represented by the coefficients: b_0, b_i, $\cdots$, b_L (Figure 18).

A related model, used by Palda [44] and others, includes previous sales as well as advertising as explanatory variables.

$$s_t = a_0 + a_1 s_{t-1} + b_0 x_t \qquad (25)$$

Meaningful values of a_1 are in (0, 1). This model also has a linear steady state response function, $s = [a_0/(1 - a_1)] + [b_0/(1 - a_1)]x$, and an exponential (geometric) impulse response with nth term $b_0 a_1{}^n$.

The two models differ considerably in statistical estimation properties, a fact which has generated considerable discussion (Houston and Weiss [24]) but from our point of view they are similar, since model (25) can be put into the form of (24) with $L = \infty$ by successive substitutions. We note that either model can be cast into the Brandaid format of (18).

These models contain very few of the advertising phenomena described earlier. Linear response is not credible over an indefinite range and obviously fails the requirements of P2.

The impulse response of (24) is versatile but rise times and decay times between steady state levels are essentially the same. To see this, observe that, if sales are in steady state under advertising rate x and we increase the rate by Δ, then n periods later sales will be incremented by $\Delta(\sum_{i=0}^{n} b_i)$. If, after establishing steady state at the new higher advertising, we decrease advertising by Δ back to x, then n periods later sales will be reduced by $\Delta(\sum_{i=0}^{n} b_i)$, the same amount. This is sketched in Figure 19. Thus linear models fail phenomenon P1.

Linear models have been extended to include competitive advertising variables. (See, for example, Picconi and Olson [49] Model 5.) This is desirable but, of course, does not circumvent the difficulties already discussed. We also note that Box-Jenkins model building techniques,

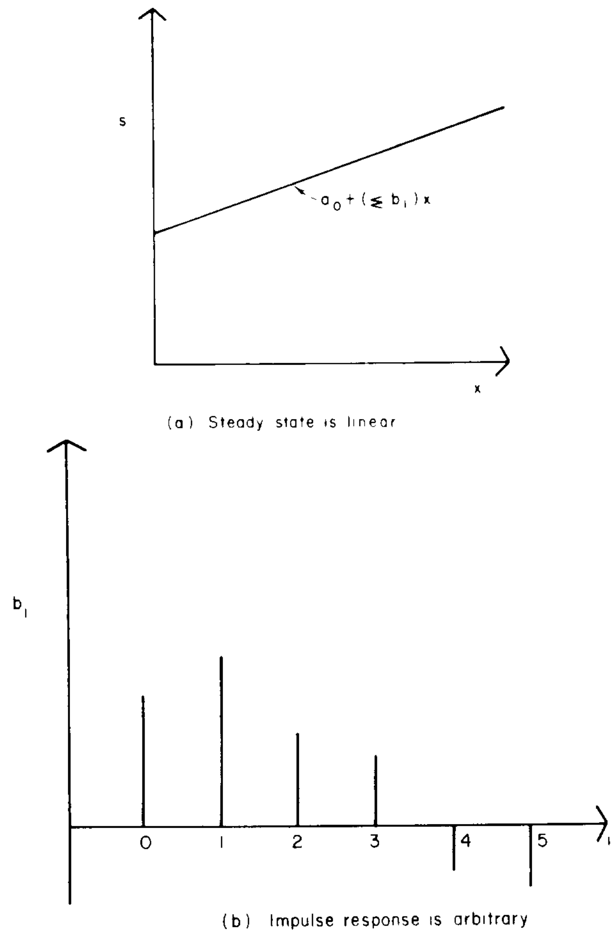


Figure 18. Steady state and impulse response of linear model (24).

although quite different in approach from standard econometric methods, produce models that are linear, or possibly product form (Helmer and Johansson [22]).

Product form Models. Many writers use models of the form

$$s_t = a_0 \Pi_{i=0}^{L} x_{t-i}^{b_i} \qquad (26a)$$

which, after taking logs, becomes linear in the constants:

$$\ln s_t = \ln a_0 + \textstyle\sum_{i=0}^{L} b_i \ln x_{t-i}. \tag{26b}$$

A lagged sales term may be added:

$$s_t = a_0 s_{t-1}^{a_1} \textstyle\prod_{i=0}^{L} x_{t-i}^{b_i} \tag{27}$$

and sometimes more than one. Logs again linearize the expression with respect to the constants and thereby greatly simplify the task of estimating them from data. Models (26) and (27) are analogs of the linear (24) and (25). The product form is widely used. Examples may be found, for instance, in Montgomery and Silk [41] and many in Lambin [30].

Product form models have an obvious defect, namely, zero advertising produces zero sales and, if lagged advertising terms are included, zero advertising in any lagged period produces zero sales in the current period. The situation is particularly acute for applications with short period

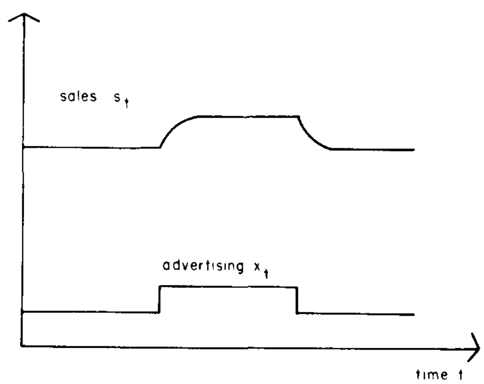


Figure 19. In the linear model (24) and product form model (26) a rapid rise time also means a rapid decay time.

lengths (e.g., months or weeks), since zero advertising in such intervals is quite common. A constant can be added to the advertising variable but an a priori constant represents a strong assumption about the shape of the response function and letting the calibration pick the constant loses the advantages of linearity for estimation (Naert and Weverbergh [42]).

Models in product form fail to conform to our required phenomena in other ways. S-shaped response is precluded. Rise and decay from steady state involve symmetric factors. Thus in (26) if sales are in steady state with advertising x and a jump of Δ is made in advertising, then n periods later sales will be multiplied by a factor $(1 + (\Delta/x))^K$ where $K = \sum_{i=0}^{n} b_i$. If, after reaching steady state with advertising of $x + \Delta$, advertising is reduced to x, then n periods later sales will be divided by the same factor.

Thus we conclude that the usual product form models fail to exhibit the key phenomena P1 and P2.

Models Additive in Nonlinear Functions of Advertising. A number of writers (e.g. Lambin [29]), have used models like (24) with the change that x_t is the share of advertising, i.e., x_t = (brand advertising in t)/(sum of advertising of all brands in t). Often this is coupled with s_t changed to market share. A model of this type satisfies two important goals: it is nonlinear in brand advertising and contains competitive effects. However, the simple share approach does not permit competitors to have different effectiveness and is rather rigid in its nonlinearity. For example, it cannot be S-shaped and so fails P2.

A variant is to use relative advertising, i.e., the denominator of x_t excludes the brand's own advertising (e.g. Clarke [12]). Also product forms are sometimes used. However, the drawbacks cited above remain. In other cases (Palda [44], Picconi and Olson [49]), equation (25) is used with x_t equal the log of advertising in t. This produces diminishing returns but cannot be S-shaped, has symmetrical rise and decay times, and loses meaning at zero advertising.

Simultaneous Equation Models. A serious problem arises in analyzing historical data because many companies set their advertising budgets, at least in part, on the basis of sales. If the direction of causality between advertising and sales is partly reversed, biased and spurious results can occur (Schmalensee [56]). The problem is potentially quite acute in annual time series data, but such data may have irreparable difficulties anyway in the form of aggregation biases, as discussed by Clarke [11]. Montgomery and Silk [41] argue that simultaneity problems are slight for time periods of, say, a month because companies do not react to competitive advertising on a month-to-month basis. Silk, in private communication, suggests that a more serious simultaneity problem arises in cross-sectional studies because of companies' geographic allocation rules.

In any case, simultaneous equation methods have been brought to bear on sales-advertising time series. Bass [2] and Bass and Parsons [4], for example, use the technique. However, the advertising response models generally used in the equation systems are product form. As a result they have the problems already discussed.

What can we conclude? First, most of the commonly used econometric models of sales response to advertising do not have structures that will accommodate the set of the dynamic phenomena identified earlier. These models are particularly weak in flexibility of shape for the response curve and in allowing different rise and decay rates. None of the models consider phenomenon, P5: sales increases under increased advertising decay with constant advertising. However, a model-builder would be unlikely to hypothesize this phenomenon without experimental evidence like that provided by Haley.

To this writer the standard econometric forms (24–27) are not so much models of advertising as convenient functions fit to the advertising response process in the neighborhood of historical operations. Such a fitting process may be useful. For example a linear model might well be reasonable if the data do not contain a large enough variance in advertising to permit meaningful calibration of a nonlinear model. The coefficient from a linear statistical model might be combined with estimates from other sources about the effects of very large or very small advertising rates to calibrate a decision model. However, the purpose of building the statistical model would then be quite different from our objectives here, which are to find the structure of advertising response that might appropriately be incorporated into the decision model.

The sheer volume of econometric work has led to some empirical generalizations. For example Clarke [11] makes a convincing case for a short term effect of advertising on the order of a few months. He also challenges certain empirically based arguments for long run effects by arguing that they are artifacts of the time period used in the econometric work. Lambin [30] also draws generalizations from his massive study, although some are not entirely persuasive. For example, he says that there is no S-curve because product form and logarithmic models fit better than linear ones. This seems an insufficient argument and, indeed, he seems to contradict himself by later advocating the existence of threshold effects.

A Priori Models with Calibration

A number of researchers have taken the approach of defining advertising models rather independently of standard econometric forms and then devising means to calibrate them on specific historical data bases. This is an important direction of research, although, as with all nonexperimental data, the researcher is dependent on historical variations to make measurement possible. Furthermore, most of the more elaborate models are nonlinear in some of the parameters. This introduces a host of calibration problems, not the least of which is the assessment of the quality of the parameter estimates.

Kuehn, McGuire and Weiss [28] present an early and ambitious example of an a priori model calibrated on historical data. Let s_{it} = market share of brand i in time period t, p_{it} = price of brand i in t, x_{it} = advertising spending of brand i in t and a_{it} = effective advertising in t.

Unknown constants are: α, β = carryover constants for sales and advertising, b = weighting constant reflecting amount of sales not affected by advertising, ϵ = advertising sensitivity exponent, δ = price sensitivity exponent, e_i = brand i advertising effectiveness coefficient and k_i = brand i effectiveness coefficient due to other marketing activities.

$$s_{it} = \alpha s_{i,t-1} + (1-\alpha)\{bk_i p_{it}^{-\delta}/\textstyle\sum_j k_j p_{jt}^{-\delta} + (1-b)k_i p_{it}^{-\delta} a_{it}^{\epsilon}/\textstyle\sum_j k_j p_{jt}^{-\delta} a_{jt}^{\epsilon}\} \tag{28a}$$

$$a_{it} = \beta a_{i,t-1} + (1-\beta)e_i x_{it} \tag{28b}$$

$$\textstyle\sum s_{jt} = 1. \tag{28c}$$

By means of nonlinear estimation on historical time series the authors determine twelve constants required in their particular case.

The model has several interesting features. Its general form is that of (18), the Brandaid advertising submodel, but with price effects imbedded in it. The steady state response function is in the braces { } and is essentially the steady state of a Lanchester model with an additive term representing sales at zero advertising. Response can be either S-shaped or concave. It is interesting to note that the fitted value of ϵ was 2.57 so that response is S-shaped in the authors' application. Effective advertising is an exponentially smoothed function of spending (28b). The constraint (28c) forces the market shares to add to one in the model and is an integral part of the estimation. The model contains many, although not all, of the phenomena laid out earlier as desirable.

Horsky [23] builds an interesting model and calibrates it on cigarette data. He considers a two competitor case, one competitor being the brand of interest and the other the rest of the industry. Let s_{it} = market share of Competitor i in period t, x_{it} = advertising spending of Competitor i in t, a_{it} = effective advertising or goodwill of Competitor i in t, β_i = carryover constant for advertising and ρ_i = effectiveness constant for advertising.

Horsky's model for Competitor 1 is $s_{1t} - s_{1,t-1} = \rho_1 a_{1t} s_{2t} - \rho_2 a_{2t} s_{1t}$ with a symmetric equation for Competitor 2. Effective advertising is given by $a_{it} = \beta_i a_{i,t-1} + (1-\beta_i)x_{it}$, $i = 1, 2$. In our terminology this is a two-competitor Lanchester model in discrete time driven by exponentially weighted past advertising. It can have different rise and decay rates, thereby satisfying phenomenon P1. The steady state response is somewhat inflexible, being concave and having zero sales at zero advertising. Nevertheless, the model is a considerable step up in complexity from most current econometric models and nonlinear estimation is required.

Parsons [46] tackles the problem of time varying advertising effectiveness. Armed with sales and advertising data for a household cleaner from 1886 to 1905 he adds a time varying coefficient to a standard product-form econometric model and finds the change in advertising effectiveness over the product life cycle. Again, nonlinear estimation is required. Pekelman and Tse [48] model copy wearout and replacement as a time-varying coefficient in a Lanchester-like competitive model and track the coefficient with Kalman filter techniques. Turner and Wiginton [69] use nonlinear techniques to calibrate the Vidale-Wolfe model on aggregate industry sales and advertising for filter cigarettes.

These examples show that, when researchers abandon the estimation conveniences of standard econometric models, they can build more realistic models and calibrate them using nonlinear methods.

4. CONCLUSIONS

We have reviewed a large amount of material on the sales effects of advertising for established products. What can we now say about representing these processes with models?

A first conclusion is that advertising is rich with phenomena. We are dealing with communication and its influence on purchase behavior. Perhaps it is presumptuous to expect a regularity that can be reduced to models with only a few parameters. Yet measurements have brought out many recurrent characteristics: an upward response of sales that takes place soon after increased advertising; a relatively slower sales decay on withdrawal that we attribute to customer satisfaction; sales saturation at high advertising levels; a possible threshold-like effect at low levels; a change of effectiveness over time because of media and copy changes; a loss of sales due to competitive advertising; and an effect reported by Haley that an advertising increase sometimes brings only a temporary sales increase. The magnitude and timing of all these effects are of great practical interest in making advertising decisions.

At the same time many other effects remain to be discovered and understood. The S-shaped curve is still on shaky ground. Is pulsing an effective policy and, if so, how long should pulses last? Does the S-shaped curve (essentially a static notion) provide an adequate theory for deriving optimal pulsing policies? What about the reported phenomenon that advertising is more effective when sales are increasing? More measurement and understanding are called for.

A second conclusion is that, although we have an apparent richness of models, many of them are rearrangements of a few key ideas. The Vidale-Wolfe constructs are surviving well, even though generalizations of the original model are very much in order. The competitive Lanchester generalization in which advertising rate is raised to a power looks quite versatile at the moment. It needs a change that will permit positive sales at zero advertising but this could be achieved by defining a component of sales not affected by advertising. The Lanchester model can be used in differential equation form or put in discrete time, in which case it would be compatible with the Brandaid advertising submodel.

We have introduced copy and media effectiveness as a multiplier on spending, and have pointed out that extensions are needed to let copy affect total sales potential. Nowhere have we presented a model for phenomenon, P5, the temporary increase in sales under a permanent increase in advertising. A parsimonious adaptation of a new trier model might help here.

A third conclusion, and possibly a controversial one, is that the commonly used econometric models are of limited value in advertising. Their functional forms generally fail to represent advertising processes except possibly over a limited range. Add to this the problems of collinearity, autocorrelation and simultaneity; an approach that initially appeared easy for learning about advertising by applying standard tools to widely available historical data begins to look more difficult. We should specify more realistic a priori models and calibrate them using the nonlinear and robust estimation tools that are now becoming generally available.

Fourth, we observe that, at least in the literature, there is an underuse of separate calibrations for different parts of a model. Particularly for decision making, we must include in our models all the phenomena that affect the decision. This will often lead to calibrating the model in several parts from eclectic data sources.

Looking ahead, new developments in measurement offer the possibility of resolving some of the outstanding modeling issues. The potential of field experiments is by no means exhausted, although field experiments and historical analysis operating on aggregate data are likely to remain fairly blunt instruments and the greatest opportunities may lie in collecting information at the individual level. Laboratory measurements analogous to those done by Silk and Urban [63] on new product communication are likely to be helpful. Even more promising is the data spinoff of point-of-sale equipment in retail stores. Such equipment can record individual purchases in machine-readable form. Optical scanning of the Universal Product Code for grocery products is one example, and analogous equipment is penetrating department stores. The sheer quantity of data will challenge our computing and analytic resources, but therein lies the potential. The coupling of individual purchase information with observations of media exposure should permit ongoing response measurements of the type pioneered by McDonald [38] and discussed earlier. Individual level measurements also seem required to examine hypotheses being generated from behavioral science (Sawyer and Ward [55]).

At the same time, the measurements must be tied into models. Some of the current structures for describing individual choice behavior appear to be extendable to consider advertising. We note that logit models (e.g. McFadden [39]) are rather similar in form to the steady state of a Lanchester model. Maximum likelihood calibration is available. Hauser [21] provides model testing methods. In another needed direction Givon and Horsky [17] have started to study issues that arise in going from disaggregate to aggregate models.

Much remains to be done. In the next 5 to 10 years there will be abundant opportunities for understanding advertising processes better and putting this knowledge to work in improving marketing productivity.

ACKNOWLEDGMENTS

The author thanks Gary Lilien, Leonard Lodish, Richard Schmalensee, Hermann Simon and Alvin Silk for many helpful comments.

REFERENCES

1. R. L. Ackoff and J. R. Emshoff, "Advertising Research at Anheuser-Busch, Inc. (1963–68)," *Sloan Management Rev.* **16,** 1–16 (1975).
2. F. M. Bass, "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *J. Marketing Res.* **6,** 291–300 (1969).
3. F. M. Bass and D. G. Clarke, "Testing Distributed Lag Models of Advertising Effect," *J. Marketing Res.* **9,** 298–308 (1972).
4. F. M. Bass and L. J. Parsons, "Simultaneous-Equation Regression Analysis of Sales and Advertising," *Appl. Econ.* **1,** 103–124 (1969).
5. J. C. Becknell, Jr., and R. W. McIsaac, "Test Marketing Cookware Coated with Teflon," *J. Advertising Res.* **3,** 2–8 (1963).
6. D. E. Bell, R. L. Keeney and J. D. C. Little, "A Market Share Theorem," *J. Marketing Res.* **12,** 136–141 (1975).
7. B. Benjamin, W. P. Jolly and J. Maitland, "Operational Research and Advertising; Theories of Response," *Opnl. Res. Quart.* **11,** 205–218 (1960).
8. B. Benjamin and J. Maitland, "Operational Research and Advertising: Some Experiments in the Use of Analogies," *Opnl. Res. Quart.* **9,** 207–217 (1958).
9. D. Bloom, A. Jay and T. Twyman, "The Validity of Advertising Pretests," *J. Advertising Res.* **17,** 7–16 (1977).
10. C. Burdet and S. P. Sethi, "On the Maximum Principle for a Class of Discrete Dynamical Systems with Lags," *J. Optimization Theory Application* **19,** 445–453 (1976).
11. D. G. Clarke, "Econometric Measurement of the Duration of Advertising Effect on Sales," *J. Marketing Res.* **18,** 345–357 (1976).
12. D. G. Clarke, "Sales-Advertising Cross-Elasticities and Advertising Competition," *J. Marketing Res.* **10,** 250–261 (1973).
13. K. R. Deal and S. Zionts, "A Differential Game Solution to the Problem of Determining the Optimal Timing of Advertising Expenditures," *Proceedings of the Second Annual Northeast Regional AIDS Conference*, American Institute of Decision Sciences, Atlanta, Ga., 1973.
14. G. J. Eskin and P. H. Barron, "Effects of Price and Advertising in Test-Market Experiments," *J. Marketing Res.* **14,** 499–500 (1977).
15. L. Friedman, "Game Theory in the Allocation of Advertising Expenditures," *Opns. Res.* **6,** 699–709 (1958).
16. D. H. Gensch, "A Computer Simulation Model for Selecting Advertising Schedules," *J. Marketing Res.* **6,** 203–214 (1969).
17. M. Givon and D. Horsky, "Market Share Models as Approximators of

Aggregated Heterogeneous Brand Choice Behavior," *Management Sci.* **24,** 1404–1416 (1978).

18. J. P. GOULD, "Diffusion Processes and Optimal Advertising Policy," in *Microeconomic Foundations of Employment and Inflation Theory*, E. S. Phelps et al. (eds.), W. W. Norton, New York, 1970.
19. S. K. GUPTA AND K. S. KRISHNAN, "Differential Equation Approach to Marketing," *Opns. Res.* **15,** 1030–1039 (1967).
20. R. I. HALEY, "Sales Effects of Media Weight," *J. Advertising Res.* **18,** 9–18 (1978).
21. J. R. HAUSER, "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach," *Opns. Res.* **26,** 406–421 (1978).
22. R. M. HELMER AND J. K. JOHANSSON, "An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising-Sales Relationship," *J. Marketing Res.* **14,** 227–239 (1977).
23. D. HORSKY, "An Empirical Analysis of the Optimal Advertising Policy," *Management Sci.* **23,** 1037–1049 (1977).
24. F. S. HOUSTON AND D. L. WEISS, "Cumulative Advertising Effects: The Role of Serial Correlation," *Decision Sci.* **6,** 471–481 (1975).
25. G. E. KIMBALL, "Some Industrial Applications of Military Operations Research Methods," *Opns. Res.* **5,** 201–204 (1957).
26. H. E. KRUGMAN, "Why Three Exposures May Be Enough," *J. Advertising Res.* **12,** 11–14 (1972).
27. A. A. KUEHN, "A Model for Budgeting Advertising," in *Mathematical Models and Methods in Marketing*, F. M. Bass et al. (eds.), Richard D. Irwin, Homewood, Ill., 1961.
28. A. A. KUEHN, T. W. MCGUIRE AND D. L. WEISS, "Measuring the Effectiveness of Advertising," in *Science, Technology and Marketing*, R. M. Haas (ed.), American Marketing Association, Chicago, 1966.
29. J.-J. LAMBIN, "A Computer Online Marketing Mix Model," *J. Marketing Res.* **9,** 119–126 (1972).
30. J.-J. LAMBIN, *Advertising, Competition and Market Conduct in Oligopoly Over Time*, North Holland, Amsterdam, 1976.
31. J.-J. LAMBIN, P. A. NAERT AND A. BULTEZ, "Optimal Marketing Behavior in Oligopoly," *European Econ. Rev.* **6,** 105–128 (1975).
32. G. L. LILIEN, "Model Relativism: A Situational Approach to Model Building," *Interfaces* **5,** 11–18 (1975).
33. J. D. C. LITTLE, "A Model of Adaptive Control of Promotional Spending," *Opns. Res.* **14,** 1075–1097 (1966).
34. J. D. C. LITTLE, "Brandaid: A Marketing-Mix Model, Parts 1 and 2," *Opns. Res.* **23,** 628–673 (1975).
35. J. D. C. LITTLE AND L. M. LODISH, "A Media Planning Calculus," *Opns. Res.* **17,** 1–35 (1969).
36. L. M. LODISH, "Empirical Studies on Industrial Response to Exposure Patterns," *J. Marketing Res.* **8,** 212–218 (1971).
37. D. H. MANN, "Optimal Theoretic Stock Models; a Generalization of Delayed Response from Promotional Expenditures," *Management Sci.* **21,** 823–832 (1975).

38. C. McDonald, "What is the Short-Term Effect of Advertising?" *Proceedings of 1970 ESOMAR Conference*, Barcelona, 1970.
39. D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, P. Zarembka (ed.), Academic Press, New York, 1970.
40. H. D. Mills, "A Study in Promotional Competition," in *Mathematical Models and Methods in Marketing*, F. M. Bass et al. (eds.), Richard D. Irwin, Homewood, Ill., 1961.
41. D. B. Montgomery and A. J. Silk, "Estimating Dynamic Effects of Marketing Communications Expenditures," *Management Sci.* **18,** B485–501 (1972).
42. P. A. Naert and M. Weverbergh, "Multiplicative Models with Zero Values in the Explanatory Variables," working paper, University of Antwerp, 1978.
43. M. Nerlove and K. J. Arrow, "Optimal Advertising Policy under Dynamic Conditions," *Economica* **29,** 129–142 (1962).
44. K. Palda, *The Measurement of Cumulative Advertising Effects*, Prentice-Hall, Englewood Cliffs, N. J., 1964.
45. L. J. Parsons, "A Rachet Model of Advertising Carryover Effects," *J. Marketing Res.* **13,** 76–79 (1976).
46. L. J. Parsons, "The Product Life Cycle and Time-varying Advertising Elasticities," *J. Marketing Res.* **12,** 476–480 (1975).
47. L. J. Parsons and R. L. Schultz, *Marketing Models and Econometric Research*, North Holland, Amsterdam, 1976.
48. D. Pekelman and E. Tse, "Experimentation and Control in Advertising, an Adaptive Control Approach," Wharton School Working Paper 76-04-01, University of Pennsylvania, Philadelphia, 1976.
49. M. J. Picconi and C. L. Olsen, "Advertising Decision Rules in a Multibrand Environment; Optimal Control Theory and Evidence," *J. Marketing Res.* **15,** 82–92 (1978).
50. A. G. Rao, Private Communication, March 1979.
51. A. G. Rao, "Productivity of the Marketing-Mix; Measuring the Impact of Advertising and Consumer and Trade Promotions on Sales," paper presented at ANA Advertising Research Workshop, New York, 1978.
52. A. G. Rao, *Quantitative Theories in Advertising*, John Wiley & Sons, New York, 1970.
53. A. G. Rao and P. B. Miller, "Advertising/Sales Response Functions," *J. Advertising Res.* **15,** 7–15 (1975).
54. M. W. Sasieni, "Optimal Advertising Expenditures," *Management Sci.* **18,** 64–72 (1971).
55. A. Sawyer and S. Ward, "Carryover Effects in Advertising Communication; Evidence and Hypotheses from Behavioral Science," in *Cumulative Advertising Effects: Sources and Implications*, D. G. Clarke (ed.), Report 77-111 Marketing Science Institute, Cambridge, Mass., 1977.
56. R. Schmalensee, *The Economics of Advertising*, North Holland, Amsterdam, 1972.
57. R. Schmalensee, "A Model of Advertising and Product Quality," *J. Political Econ.* **86,** 485–503 (1978).

58. S. P. SETHI, "A Comparison between the Effect of Pulse versus Continuous Television Advertising on Buyer Behavior," in *1971 Combined Proceedings*, F. C. Allvine (ed.), American Marketing Association, Chicago, 1971.
59. S. P. SETHI, "Dynamic Optimal Control Models in Advertising; a Survey," *SIAM Rev.* **19,** 685–725 (1977).
60. S. P. SETHI, "Optimal Control of a Logarithmic Advertising Model," *Opnl. Res. Quart.* **26,** 317–319 (1975).
61. S. P. SETHI, R. E. TURNER AND C. P. NEUMAN, "Policy Implications of an Intertemporal Analysis of Advertising Budgeting Models," in *Proceedings 4th Annual Midwest AIDS Conference*, American Institute of Decision Sciences, Atlanta, Ga., 1973.
62. M. F. SHAKUN, "A Dynamic Model for Competitive Marketing in Coupled Markets," *Management Sci.* **12,** B525–530 (1966).
63. A. J. SILK AND G. L. URBAN, "Pretest Market Evaluation of New Packaged Goods; a Model and Measurement Methodology," *J. Marketing Res.* **15,** 171–191 (1978).
64. J. L. SIMON, "New Evidence for No Effect of Scale in Advertising," *J. Advertising Res.* **9,** 38–41 (1969).
65. M. K. STARR, "OR in Marketing; a Case of Failure and Renewal," in *OR'78*, K. B. Haley (ed.), North Holland, Amsterdam, 1978.
66. E. C. STRONG, "The Use of Field Experimental Observations in Estimating Advertising Recall," *J. Marketing Res.* **11,** 369–378 (1974).
67. W. R. SWINYARD AND M. L. RAY, "Advertising-Selling Interactions; an Attribution Theory Experiment," *J. Marketing Res.* **14,** 509–516 (1977).
68. C. S. TAPIERO, "On-Line and Adaptive Optimum Advertising Control by a Diffusion Approximation," *Opns. Res.* **23,** 890–907 (1975).
69. R. E. TURNER AND J. C. WIGINTON, "Advertising Expenditure Trajectories; an Empirical Study for Filter Cigarettes 1953–1965, *Decision Sci.* **7,** 496–509 (1976).
70. G. L. URBAN, "Building Models for Decision-Makers," *Interfaces*, **4,** 1–11 (1974).
71. M. L. VIDALE AND H. B. WOLFE, "An Operations Research Study of Sales Response to Advertising," *Opns. Res.* **5,** 370–381 (1957).
72. R. S. WEINBERG, *An Analytical Approach to Advertising Expenditure Strategy*, Association of National Advertisers, New York, 1960.
73. A. R. WILDT, "Estimating Seasonal Marketing Response using Dummy Variables," *J. Marketing Res.* **14,** 34–41 (1977).
74. D. R. WITTINK, "Exploring Territorial Differences in the Relationship between Marketing Variables," *J. Marketing Res.* **14,** 145–155 (1977).
75. F. S. ZUFRYDEN, "Media Scheduling; a Stochastic Dynamic Model Approach," *Management Sci.* **18,** B145–154 (1973).