Chapter 9

Econometric and Time-Series Market Response Models*

Dominique M. Hanssens
Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, CA 90024-1481, USA

Leonard J. Parsons
School of Management, Georgia Institute of Technology, Atlanta, GA 30332-0520, USA

1. Introduction

Marketing managers need to know how their markets will respond, both in the short run and in the long run, to the actions they take. The heart of marketing management practice then is understanding the market mechanism governing a particular market. The key component of such a mechanism is a sales-response function, how unit sales are affected by the controllable actions of the firm, the uncontrollable activities of competitors, and autonomous events in the environment. When a firm wants to describe a more complete model of a market mechanism, multiple relations may arise. Not only will there be a sales-response function, but there also may be competitive reaction functions, vertical market structures, cost functions, and other behavioral relations. A market mechanism specifies the connections among these relations as well as among individual variables.

Various elements constitute a market-response model. In any given situation, the model builder may or may not have advance knowledge of these factors. For example, the analyst may be able to identify the elements of the marketing mix, but know very little about the functional form of the model. It is the task of careful empirical analysis, using econometric and time-series methods, to advance the model builder to a higher state of knowledge. Econometrics is the application of statistical and mathematical techniques for the measurement (estimation and testing) of structural relations in predominantly nonexperimental situations. The major technique used in econometrics is regression analysis and its extensions.

*This chapter is a condensation and update of Hanssens, Parsons & Schultz [1990]. We would like to acknowledge the contributions of our collaborator in that original work, Randall L. Schultz of the University of Iowa. We thank Albert Bemmaor (ESSEC), Peter Leeftang (Groningen), Gary Russell (Toronto) and reviewers for their thoughtful comments.
Time-series analysis is the application of one of a wide range of techniques, including univariate and multivariate time series analysis, to examine data that evolve over time. The underlying observations are usually obtained at regular intervals.

Our knowledge about the modeling environment can be organized in the following way:

- level 0: only the information set (i.e., a collection of relevant variables) is known;
- level 1: the causal ordering among variables is known; and
- level 2: the functional form, causal ordering and lag structure are known.

Econometric-time series techniques (ETS) are not appropriate to situations with less than level-0 knowledge. We must start with an information set developed from subject-matter theory or directly from managers. For example, the concept of the marketing mix leads to a commonly used information set in market-response modeling consisting of product sales, price, distribution, sales force, and communication efforts. Once a level-0 prior knowledge is obtained, econometric-time series methods can make substantial contributions in moving the marketing scientist up the knowledge hierarchy.

At level 0, empirical methods should be used to establish the direction of causality among marketing variables. This can only be accomplished with time-series data because it requires the temporal ordering of events. At level 1, the model builder is not ready to estimate parameters or predict sales. The functional form and dynamic structure of the model must first be specified. The latter may require a different set of techniques involving time-series analysis—univariate techniques for lag structure specification and multiple time-series methods. At level 2, the model builder estimates the parameters of a fully specified model using econometric techniques. The analyst may verify the adequacy of the model via testing procedures. The model may then be used for forecasting and/or marketing planning.

We will first look at how we might model the market mechanism from an econometric point of view. Our perspective is primarily, but not exclusively, static in nature; that is, actions and responses take place in the same time interval. We will examine how to describe how different marketing phenomena may be captured with specific functional forms. These functional forms, of course, may also be used in dynamic models. By initially only discussing static models, we are simplifying our discussion. (Space limitations preclude us from addressing the numerous practical problems such as multicollinearity that plague applied researchers trying to implement market-response models.) We will then turn to extending this static analysis to capture the dynamics of the process, especially by gaining insights from time-series analysis. Again to simplify our discussion, we will focus on linear time-series models. A market mechanism, however, may be both dynamic and nonlinear. The dynamic structure of marketing variables themselves will first be addressed, to be followed by discussions of leads and lags among marketing variables and the assessment of the direction of causality. Dynamic properties of sales-response functions will be discussed in more detail. Marketing generalizations that have been uncovered as well as empirical evidence on the shape of the sales response function will be reported. We would usually now step to the normative
domain once we have a response model estimated. However, we will not go on to discuss what we do with our modeling results in this review. The interested reader is referred to ‘Improving marketing decisions’ in Hanssens, Parsons & Schultz [1990] as well as recent articles such as Levy & Simon [1987], Luhmer, Steindl, Feichtinger, Hartl & Sorger [1988a, 1988b], Sasiemi [1989], Doyle & Saunders [1990], Cool & Devinney [1992], Feinberg [1992], Mantrala, Sinha & Zoltner [1992] and Mesak [1992]. Finally, current research issues will be noted.

2. Market mechanisms

The core relation in the market mechanism is the sales-response function. The dependent variable in a sales response function should be a quantity, rather than a monetary measure, of sales. One reason for this is that we need quantity sales forecasts for planning purposes. Another is to avoid spurious correlation arising from price possibly being on both sides of the sales-response function, e.g. revenue (price \times \text{quantity sold}) = f(\text{price}). For a discussion of the complexities caused by the use of composite dependent variables such as revenue, see Farris, Parry & Alawadi [1992] and Alawadi & Farris [1992].

Sometimes a firm might find it advantageous to decompose its sales-response function into two relations: an industry demand function and a market-share model. The first relation would describe how various factors influence industry sales; the second how various factors, which may or may not include some or all of those factors affecting industry sales, influence the firm’s market share. Many firms evaluate their relative success in terms of selective demand position or market share. Three reasons for this are suggested. One is that the product category is simply mature and the primary demand has a zero growth rate; e.g. the frequently purchased, inexpensive, consumable good studied by Beckwith [1972], or the product category within the hypnotics and sedative segment of the British pharmaceutical market described by Leeftang, Mijatovich & Saunders [1992]. Another is that trends in primary demand are frequently out of the control of the firm and affect the industry as a whole. The third is that marketing instruments, in particular advertising, may have minimal impact on total industry sales. Instead, the setting of managerial decision variables serves to allocate this total amount among the competing firms. In many cases, however, a single relation between a firm’s sales and both its own actions and environmental variables will be preferred. One practical problem with market-share models is defining the relevant market, i.e. industry boundaries. Decomposition, like many design alternatives in response modeling, depends on the nature of the response problem.

A firm often must be able to forecast the levels of competitors’ marketing instruments. If the competitors make their decisions without regard to the firm’s actions, then time-series analysis might be used to project future values for their marketing activities. If the competition reacts to the firm’s actions, then reaction functions should be specified. If both the firm and its competitors are simultaneously
trying to optimize their results in light of their opponents' likely behaviors, then optimal control theory and game theory might be appropriate.

Most firms do not sell directly to their customers. Usually one or more intermediaries exist in a channel of distribution. For example, a channel for a consumer packaged good might look like this:

```
Factory ———> Chain/distributor ———> Retail ———> Customer
  shipments       warehouse  withdrawals  store  sales
```

*Each channel member has its own sales-response function.* For example, Blattberg & Levin [1987] study the effectiveness and profitability of trade promotions using a market mechanism containing two relations— one for shipments and another for consumer sales. Factory shipments respond to trade promotions. Shipments increase sharply during the interval when an allowance is given on product purchases, then fall markedly after the promotion is over. A similar pattern occurs just before a major price increase as the trade stocks up on a product at the existing price. Peaks and valleys in retail sales occur in response to the presence or absence of temporary sales promotions such as price features, special displays, and coupons. Findley & Little [1980] note that dynamics of warehouse withdrawals tend to be smoother than factory shipments and retail sales. They argue that this smoothing is mainly due to the buffer effect of retail inventories.

In many applications, a product's cost per unit, exclusive of marketing costs, is assumed to be constant. This is a satisfactory approximation in most circumstances. However, there are times when more attention should be given to the *cost function.* For instance, price promotions typically cause both consumer sales and factory shipments to be uneven. Irregular factory shipments often mean higher production and/or inventory costs. Another exception occurs in the case of technological innovations, including consumer durables. For these products, total unit costs usually decline as experience with producing product is gained [Hall & Howell, 1985]. Total unit costs also tend to decline in response to competitive entry. As new brands enter a market, the resultant growth of industry output forces price downward; and consequently, costs must be reduced if profitability is to be maintained [Devinney, 1987].

The design of response models involves variables, relations, functional forms, and data. Variables represent the building blocks of a response study. An analysis of price elasticity, for example, would require at least two variables: price and unit sales. Relations deal with the connections among variables. To answer a question about the magnitude of price elasticity, it would be necessary to examine the special relation of price to unit sales. Functional forms refer to the nature of a relation. One form of a relation between price and sales could be linear; a form such as this would give both mathematical and substantive meaning to the relation. Finally, data are the actual realizations of variables. We will focus on market-level data primarily, but note that Wittink & Porter [1991] and Wittink, Porter & Gupta [1991] recommend using store-level data (when available) rather than market-level data for estimating sales-response functions because of aggregation.
bias present in market-level results. (Their comment is consistent with Bolton's [1989] finding that 'elasticity estimates exhibit substantial variability across stores'.) Taken together, these things provide the materials for building a response model. Let us now examine the formulation and estimation of response models.

3. Static models

A firm may simply be interested in how its advertising and price affect its sales. This firm needs to know about a single relation – its sales-response function. Its sales, the dependent variable, are determined, in part, by a set of explanatory variables, advertising and price. The set of explanatory variables might be expanded to include environmental variables as well as other decision variables of the firm. The environmental variables may represent competitive actions and autonomous phenomena such as macroeconomic variables.

A relation is made concrete by specifying its functional form. A functional form should exhibit the same properties the relation is known to possess. Possible properties of a sales-response function include what happens to sales when marketing effort is zero or very large; rate of change in sales as marketing activity increases, e.g. diminishing returns to scale; threshold effects such as a minimum advertising investment; parameter variation such as might occur over different market segments; and asymmetric response such as a different response by competitors to a decrease in price or to an increase in price. Reaction functions, since they usually represent decision rules, may be less complex. In discussing static models, we cover the shape of the sales-response function, specification of market-share models, and econometric estimation in marketing.

3.1. Shape of the sales-response function

The shape of the sales-response function may exhibit increasing, constant, or decreasing returns to scale, might be different depending on whether marketing effort is increasing or decreasing, and might change in response to changes in the environment. (Many of the functional forms of the market-response function we will cover as well as other forms are discussed in Naert & Leeflang [1978, Chapter 5], Saunders [1987], and Lilien, Kotler & Moorthy [1992, Appendix C].) Constant returns to scale occur when any unit change in marketing effort generates an equal incremental change in sales. This is true of a linear model:

$$Q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

(3.1)

where \(Q\) is unit sales, \(X_i\) is the \(i\)th explanatory variable, and \(\beta_i\) is the \(i\)th parameter. Linear relations are used by Banks [1961] and Lambert [1968] to study the impact of managerial decision variables and market share and volume, respectively; by Moriarty [1985] to examine promotional effects on intra- and inter-brand performance; by Blattberg & Neslin [1990, pp. 197–204] to illustrate using regression
analysis to investigate the effects on instant coffee sales of promotion: feature, display, and price-cut activities; and by Leeflang, Mijatovic & Saunders [1992] to investigate (dynamic) promotional effects on market share.

When sales always increase with increases in marketing effort, but each additional unit of marketing effort brings less in incremental sales than the previous unit did, a sales-response curve is said to exhibit diminishing returns to scale. A functional form that meets this requirement, for certain parameter values, is the multiplicative model:

\[ Q = e^{\beta_0} X_1^{\beta_1} X_2^{\beta_2} \cdots X_k^{\beta_k}, \quad \beta_i < 1. \] (3.2)

This model has the attractive property that the power coefficient of the marketing instrument can be directly interpreted as that instrument’s short-term elasticity. Taking the natural logs of both sides makes the model linear. Bass & Parsons [1969], Lambin [1976], and many others employ multiplicative models for sales response to advertising of frequently purchased branded goods. (While increases in most marketing decision variables such as advertising expenditures, sales calls, deal discounts, and retail availability lead to higher sales, increases in price lead to lower sales. Consequently, to make it comparable with other marketing variables, price is sometimes expressed in the response function in reciprocal form – one over price. In passing, we note that deal discount is often modeled using a percentage discount from regular price (RP – DP)/RP.) Gopalakrishna & Williams [1992] use a multiplicative model in a planning and performance assessment of industrial trade shows. A special case of the multiplicative model when there is only a single instrument and \( \beta_i \) lies between 0 and 1 is the fractional-root model, which itself is known as the square-root model when the fraction equals one-half.

Another concave sales-response function is the semi-logarithmic model:

\[ Q = \beta_0 \ln X. \] (3.3)


Although sales response to most marketing variables exhibits diminishing returns to scale, sales response to decreases in price may exhibit increasing returns to scale. An exponential model,

\[ Q = Q^* e^{-\beta_0 X}, \quad \beta_0 > 0, \] (3.4)

is employed by Cowling & Cubbin [1971] to explain the United Kingdom market for cars in terms of quality-adjusted price, by Blattberg & Levin [1987] to assess the effect of trade promotions on factory shipments, by Krishnamurthi & Raj [1988] to model household purchase quantity as a function of price, by Blattberg

Sometimes a response function might be \textit{S-shaped}. More precisely, we are discussing 'nicely convex–concave' functions. Initially, sales may exhibit increasing returns to scale and then diminishing returns to higher levels of marketing effort. Bemmaor [1984] considers a \textit{log-reciprocal model} in his study of the advertising threshold effect:

\[
Q = \exp\left(\beta_0 - \frac{\beta_1}{X}\right), \quad \beta_0 > 0. \tag{3.5}
\]

Even when marketing effort is zero, a firm might still have sales due to loyal buyers, captive buyers or impulse buyers. This would be a minimum sales potential, or \textit{base sales}, \( Q_0 \). Most functional forms can be modified appropriately by simply adding a positive constant as Metwally [1980] does to a log-reciprocal model. There is also a finite achievable upper limit to sales no matter how much marketing effort is expended. Buyers become insensitive to the marketing stimuli or find themselves purchasing at their capacities or capabilities. This maximum sales potential is called \textit{saturation}, \( Q^* \). Many commonly used sales-response functions work well despite not formally modeling saturation even though it must exist. This is because firms do not find it profitable to be operating very near saturation; and so these functional forms are adequate for the actual operating range, especially for frequently purchased consumer goods. A saturation level is explicitly represented in the \textit{modified exponential model}:

\[
Q = Q^*(1 - e^{-\beta X}). \tag{3.6}
\]


\[
\ln\left(\frac{Q - Q_0}{Q^* - Q}\right) = \ln \beta_0 + \sum_{j=1}^{J} \beta_j X_j. \tag{3.7}
\]

Johansson [1973] used survey data to estimate the minimum sales level and the saturation level for a new woman's hairspray in one version of a logistic model with market share as the dependent variable. His estimate of minimum sales was the proportion of repeaters and his estimate of saturation was the trial proportion. When a priori information is not available, we must use a functional form that
allows the intercept and saturation level to be estimated. One such functional form was used by Little [1970] in his ADBUG model:

\[ Q = Q_0 + (Q^* - Q_0) \frac{X^{\beta_2}}{\beta_3 + X^{\beta_2}}. \]  

(3.8)

An additional phenomenon has been theorized – supersaturation. Supersaturation occurs when higher levels of marketing effort result in lower sales than lower levels of effort were able to achieve; that is, the marginal rate of change is negative. Perhaps this might occur when too much marketing effort causes a negative response; for example, a buyer might feel that an excessive number of visits by a salesperson is intolerable. The simplest functional form that represents this effect is the quadratic model:

\[ Q = \beta_0 + \beta_1 X - \beta_2 X^2. \]  

(3.9)

Some positive amount of marketing effort might be necessary before any sales impact can be detected. For example, the expenditure of only one thousand dollars in a highly competitive mass market is unlikely to show a sales effect. The minimum effort to show an effect is called a threshold. This might be modeled by simply adding a positive constant to a marketing instrument. The fact that the quantity sold must be nonnegative (ignoring the possibility that returned product exceeds product sales) implies some functional forms have thresholds built into them, for example, the saturation model with decreasing returns:

\[ Q = \beta_0 - \beta_1 X^{-\beta_1}. \]  

(3.10)

A special case (\( \beta_2 = 1 \)) of this, the reciprocal model, is applied by Ward [1975] to the processed-grapefruit industry.

One marketing decision variable may moderate or enhance another. A fast-food company may advertise on local television just before the drop of coupons in newspapers. The combination of variables is more effective than would be predicted from each variable separately. The power model can be extended to take into account all the interactions among marketing decision variables. A very general way for representing interactions is the transcendental logarithmic (translog) model:

\[ \ln Q = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_{12} \ln X_1 \ln X_2 + \beta_{13} \ln X_1 \ln X_3 + \beta_{23} \ln X_2 \ln X_3 + \beta_1 (\ln X_1)^2 + \beta_2 (\ln X_2)^2 + \beta_3 (\ln X_3)^2. \]  

(3.11)

The translog functional form is a quadratic approximation to any continuous function. The elasticities of the marketing instruments vary with changes in the entire marketing mix. (A constant-elasticity model is a special case.) Jagpal, Sudit
& Vinod [1982] estimate a translog model for Lydia Pinkham sales and advertising. The full interaction model can quickly become unwieldy when all the possible interactions among marketing decision variables are taken into account. Because of small sample sizes or estimation problems, this model is usually simplified by having some of the parameters equal each other, and often equal 1 or 0 as well. Prasad & Ring [1976], for example, look at four main effects plus only the six pairwise interactions. A special case of the translog model is the multiplicative nonhomogeneous model (when \( \beta_{11} = \beta_{22} = \beta_{33} = 0 \)). Jagpal, Sudit & Vinod [1979] illustrate this model with Palda’s [1964] Lydia Pinkham data. Jagpal [1981] also uses it in his investigation of cross-product media advertising effects for a commercial bank. In turn, the most popular sales-response function, the multiplicative model is a special case of the multiplicative nonhomogeneous model in which only the highest-order interaction is retained. Hanssens & Levien [1983] specify multiplicative functional forms for the relationships in their econometric model of recruitment marketing in the US Navy.

Interaction among marketing variables may also be modeled by making the parameters of one marketing variable a function of other marketing variables, e.g. the price elasticity a function of advertising. More formally, the parameter vector \( \beta \) in some models such as \( q = f(\beta, X) \) may exhibit variation and this coefficient variation may be systematic. Systematic variation implies that the parameter vector can be expressed as a function of other parameters \( \alpha \) and observable variables \( Z \). This set of variables may include some of the variables in \( X \). An application to sales-call elasticities is given in Parsons & Vanden Abeele [1981].

The deterministic relationship can be made stochastic by adding a disturbance term. When this relationship is linear and it is embedded in a linear response function, the resultant model has an heteroscedastic disturbance term. Such a model is used by Gatignon [1984] to investigate the influence of advertising on a market’s price sensitivity and by Gatignon & Hanssens [1987] to examine factors influencing sales-force effectiveness.

Our discussion of functional forms has emphasized the specification of a functional relation between two or more variables. Such a functional relation does not explain everything. After all, it is only a model incorporating the salient features of some marketing phenomenon. Moreover, only the deterministic part of the underlying mechanism has been specified. There may be some inherent randomness in the marketplace; for instance, unpredictable variety seeking by customers. Consequently, any specification of a relationship must be expanded to include a random disturbance term that captures equation error, i.e. the difference between actual observations of a dependent variable and those predicted by a particular functional relationship, i.e. add \( v \) to, or multiply \( e^v \) times, the functional form of choice as appropriate. Typically \( v = \omega \), where \( \omega \) is white noise.

In situations where the data represent observations from a combination of a cross-section and a time series, the disturbance term is sometimes partitioned into three components. The components are the individual or cross-section effect, the temporal effect, and the remaining effects which vary over both individuals and time periods. Parsons [1974] and Moriarty [1975] use variants of this variance-components formulation. Parsons studied the advertising, retail availability, and
sales of new brands of ready-to-eat cereals. Moriarty studied regional fluctuations in the effectiveness of marketing decision variables for a single brand.

Discrimination among alternative models is impossible if they are observationally equivalent. This occurs when two or more theories yield exactly the same implications about observable phenomena in all situations. Under such conditions no sample, no matter how large, can resolve the issue. We will use the term observational equivalence loosely to cover any observed space in which the models under consideration are not observationally distinguishable. Saunders [1987, pp. 27–29] applies different functional forms to same data sets and all fit quite well. He concludes:

"It is evident that good fit does not mean that an expression [functional form] gives the right shape. This is particularly true if an expression is used to fit data that only covers a small portion of the curve. In such cases it is not safe to draw conclusions about the shape that an expression gives beyond the limits of the data."

This warning should be kept in mind when implementing marketing policies based upon an estimated sales response function.

3.2. Specification of market-share models

When the dependent variable in a response function is market share, an additional complication is introduced. A desirable property of any market-share model is that it be logically consistent. This means that the model produces an estimate of a brand’s market share that lies between 0 and 1 when market share is expressed as a fraction. Furthermore, the sum of these estimated market shares for all brands in any given period must equal one. Violations of the range and/or sum constraints expose the internal inconsistencies of a model.

A model that must be logically consistent is one based upon the definition of market share. This model is called an attraction model [Bell, Keeney & Little, 1975]:

$$\text{MS}_i = \frac{Q_i}{\sum_{j=1}^N Q_j} = \frac{f_i(X, v_i; \beta_i)}{\sum_{j=1}^N f_j(X, v_j; \beta_j)}$$

(3.12)

By including competitors’ marketing instruments in the set of explanatory variables $X$ for each sales-response function, cross-competitive (market-share) effects can be modeled. When the multiplicative form is used for each individual sales response function, the model is known as the MCI (multiplicative competitive interaction) model [Cooper & Nakanishi, 1988; Houston, Kanetkar & Weiss, 1991]:

$$\text{MS}_{ij} = \frac{\beta_0 \prod_{k=1}^K X_{ik}^{\beta_k} v_{ii}}{\sum_{j=1}^N \prod_{k=1}^K X_{jk}^{\beta_k} v_{ij}}$$

(3.13)
The MCI model also assumes all brands have the same coefficients. In effect, a prototype brand rather than a specific brand is being estimated. The assumption that each firm has the same coefficient is relaxed in the differential-effects MCI model:

$$MS_u = \frac{\beta_{oi} \prod_{k=1}^{K} X_{ik}^\beta v_i}{\sum_{j=1}^{N} \beta_{oj} \prod_{k=1}^{K} X_{jk}^\beta v_j}$$

(3.14)

Carpenter, Cooper, Hanssens & Midgley (1988) point out that differential effects can arise for four reasons: (1) differences in marketing effectiveness; (2) differences in competitive vulnerability, i.e. relative positions in the market structure; (3) differences in marketing objectives; and (4) differences in timing of marketing efforts. The differential-effects MCI model is used by Urban [1969] in studying interdependencies among brands in a firm's product line, by Bultez & Naert [1975] in investigating competitive marketing activities in the market for an inexpensive European consumer durable, and by Carpenter, Cooper, Hanssens & Midgley [1988] in exploring an Australian household-product market. Vanden Abeele, Gijsbrechts & Vanhuele [1991] propose and empirically evaluate a cluster-asymmetry market-share model. It represents competitive cross-effects at the level of clusters of market contenders and does so through the attraction of the competitors rather than through the instruments which determine these attractions. A variant of this model is used by Bultez, Gijsbrechts, Naert & Vanden Abeele [1989] in their analysis of the retail assortment for the canned dog-food category of a Belgian hypermarket. Fockens, Leeﬂang & Wittink [1992] put forth and empirically test a hierarchical market-share model. It assumes that customers choose items according to brand attributes which are considered in hierarchical order. Competition between and within branches can be modeled separately allowing for heterogeneous competitive effects. Both the cluster model and the hierarchical model, unlike the differential-effects MCI model, require some type of a priori structuring of competition (with the benefit of reducing the number of parameters in the model to be estimated).

We have argued that a desirable property of market-share models is that they be logically consistent. Attraction models meet this requirement. However, they do so at a cost. They are not very parsimonious and are often 'overparameterized', that is, they may contain a large number of parameters relative to the size of the sample and variability in the data. As a result, linear or multiplicative market-share models are often used in practice despite not being logically consistent in most situations.

This state of affairs has led to a series of studies on the estimation and testing of market-share models [Naert & Weverbergh, 1981, 1985; Brodie & De Kuyver, 1983; Ghosh, Neslin & Shoemaker, 1984; Leeﬂang & Reuyl, 1984]. Most studies emphasize assessing the degree of heterogeneity in the parameters of a model. Hypotheses that some of the parameters are equal are tested and the descriptive
and predictive power of the alternative models are examined. The studies differ on whether the linear, multiplicative, or attraction model is 'best'. This suggests that the answer may well be criterion as well as product-specific.

3.3. Econometric estimation in marketing

Our focus is on finding estimates of the unstandardized parameters in a sales response function, or, more generally, in a market mechanism. This information will allow a marketing manager to evaluate the possible consequences of a marketing action. In this setting the use of standardized parameter estimates, sometimes called beta weights, is not appropriate. Wittink [1983a] demonstrates that 'beta weights are meaningless for applications involving managerial decision making' by showing that 'although beta weights allow for comparisons between predictor variables when the variables are measured in different units, they have no actionable interpretation to managers'.

To make estimation as easy as possible and to ensure that any resultant parameter estimates have desirable statistical properties, simplifying assumptions are made about the error structure. Again, discussion of such detail is omitted in this review. Rather we will simply point out who in marketing use the various possible econometric estimation methods.

There are numerous applications of econometrics to a single linear estimating equation in marketing. We are talking about an equation that is linear in its parameters. Certain models can be specified as nonlinear, but can be transformed into a linear-in-parameters estimating equation. Some examples of ordinary least squares estimation include Wierenga [1981] in relating the number of visitors in a recreational park to advertising effort, Wildt, Parker & Harris [1987] in analyzing sales contests, and Hagerty, Carman & Russell [1988] in estimating elasticities with PIMS (Profit Impact of Market Strategies) data.


Linear models are computationally easy to estimate. Unfortunately, most marketing phenomena are nonlinear. For example, a sales-response function is believed to exhibit diminishing returns to scale over most, if not all, of its range. Many nonlinear functional forms, however, can be transformed into linear ones for estimation purposes. Those that cannot be transformed must be estimated by nonlinear methods. Fortunately, more and more statistical and econometric procedures use nonlinear estimation because computation has become feasible. Marketing has lagged behind on this issue.

Transformations are often used to convert linearizable nonlinear structural models into linear estimating equations. Consider the most common sales-response function, the multiplicative (3.2), with a multiplicative disturbance. A linear estimating equation for this function can be found by taking the logarithms of both sides of the relationship. For example, Di Benedetto [1985] applies this transformation to a multiplicative dynamic-adjustment model of sales response to marketing-mix variables. A problem arises when an observation on a variable, especially the dependent variable, is zero \( \ln(0) = -\infty \). To get around this problem, many researchers add 1 to each observation \( \ln(1) = 0 \). Rao, Wind & DeSarbo [1988, p. 132], for example, recommend adding a small positive number to all entries. Young & Young [1975] recommend dropping such observations rather than arbitrarily setting the log-value of the dependent variable to be zero. Alternatively, when independent variables such as promotion are zero (i.e. there are often periods when no promotion is offered), researchers are now replacing the \( X's \) in the multiplicative model by \( e^{X's} \), and the resultant sales-response function is written as
\[ Q = e^{\beta_0 (e^x)^x + \beta_1 X_1} = \exp\left( \beta_0 + \sum_{i=1}^{k} \beta_i X_i \right). \] (3.15)

This is a generalized exponential model, cf. (3.4). Applications include Cooper & Nakanishi [1988], Vanden Abeele, Gisbrechts & Vanhuele [1990] and Foekens, Leeflang & Wittink [1992].

Some functional forms are intrinsically nonlinear. For instance, if the multiplicative response-function has an additive error instead of a multiplicative error, the relationship cannot be transformed and its parameters must be found by nonlinear estimation. The least squares principle can be applied to nonlinear models although the computations will be more complex. Nonlinear least squares (NLS) in general provides biased estimates of the parameter vector \( \beta \). A more serious problem is that the distribution of \( \beta \) is usually unknown even if the distribution of the disturbance term is known. We must rely on asymptotic results arising from approximations we have to make. Under suitable conditions, the NLS estimator is consistent and asymptotically normally distributed. When the error term follows the standard normal distribution, the maximum-likelihood estimator is the same as the least squares estimator as is the case for linear models. Nonlinear regression is used by Horsky [1977] to estimate market-share response to advertising in the cigarette industry, by Metwally [1980] to estimate sales response to advertising of eight Australian products, by Parker [1992] to investigate price-elasticity dynamics over the adoption life-cycle, and by Gopalakrishna & Chatterjee [1992] to study the impact of the industrial communications mix of a firm marketing electrical cables.

The coefficients in the standard linear model are assumed to be constant. This may well not be true if micro-units in a cross-section study respond differently or if the environment changes over time. Thus we have nonconstant coefficients. If the changes are systematic and deterministic, OLS regression can be used. However, if the systematic changes also incorporate a random disturbance, then estimated (a.k.a. approximate or feasible) generalized least squares (EGLS) must be used because the error term will be heteroscedastic. Gatignon [1984] extends this approach to take into account constraints on some of the parameters in his model.

4. Dynamic models

Marketing for a firm rarely takes place in a static environment. Customers and competitors anticipate or react to the firm’s actions. Their adjustment processes are one basis for believing market mechanisms should be dynamic. We will discuss, in order, three major scenarios in market-response modeling for which a dynamic approach is appropriate. In each case we will discuss the modeling issues, make a brief reference to estimation issues, and conclude with some applications and findings in the marketing literature. A much more extensive treatment of these issues may be found in HPS [Hanssens, Parsons & Schultz, 1990].
4.1. The dynamic structure of marketing variables by themselves

There are a number of scenarios in model-based planning and forecasting which make it desirable to analyze a marketing time-series strictly as a function of its own past. These scenarios can be organized in three categories:

1. We have developed a planning model relating, say, product prices to product sales. However, price may be determined partially from market factors outside the control of the firm, so it is necessary to forecast prices separately. These predictions are then used to obtain sales estimates. We refer to such situations as 'forecasting exogenous variables'.

2. Our product line is so large that building individual planning models for each product is prohibitive. Nevertheless, separate forecasts for each are needed. Perhaps the company will invest in a comprehensive marketing-mix model for the four or five leading products and use extrapolative methods to handle the remaining two hundred or so items. This would be an example of 'forecasting performance variables'.

3. Sometimes it is useful to decompose a marketing time-series, say price, into a systematic, predictable part and a random, unpredictable part. For example, product sales may react differently to predictable price changes, say those due to inflation adjustments, than to unpredictable price 'shocks', say surprise deals offered by the manufacturer. This decomposition of the price variable produces a smooth, predictable price series and a residual price series which is uncorrelated over time (white noise). This is an example of 'prewhitening' a marketing time-series.

The three scenarios apply only when marketing data over time are available. Furthermore, we will assume that the data are collected in regular intervals (e.g. weekly or quarterly) and that they are sufficiently long for statistical modeling (e.g. a minimum of 30 uninterrupted observations). Under these assumptions we can apply principles of univariate time-series analysis in order to obtain extrapolative forecasts.

The time-series analyst examines the behavior of data over time as a function of deterministic and stochastic elements:

1. Deterministic elements whose outcome is perfectly predictable at any point of time. An example is the linear trend model on an arbitrary variable $Z$:

$$Z_t = \beta_0 + \beta_1 t,$$

where $t$ is a time counter. For every new period, a fixed value $\beta_1$ is added to the base level $\beta_0$ of the time series.

2. Random or stochastic components whose effect cannot be predicted with certainty. If the random component is correlated over time, it may contribute to forecasting, although imperfectly. It is referred to as a systematic time-series component. If it is uncorrelated, however, it is strictly an error term and of no use for forecasting. These terms are known as 'white noise', 'shocks', or 'innovations'.
For example:

\[ Z_t = \beta Z_{t-1} + \omega_t, \]  

(4.2)

where \( E(\omega_t) = 0, \ E(\omega_t^2) = \sigma^2 \) and \( E(\omega_t \omega_{t-k}) = 0 \) for all \( k \neq 0 \).

The first right-hand term is a systematic effect of the last period on the current period and is useful for forecasting. The second term is white noise and, while it may affect the current \( Z \) significantly, it cannot be used for estimating future \( Z \).

In conclusion, situations exist in model-based planning and forecasting where strictly extrapolative predictions of marketing variables are needed. Such models are developed using principles of modern time-series analysis, which pay particular attention to the deterministic, systematic, and pure random elements in marketing data. The most popular of these techniques is due to Box & Jenkins. In essence, this method finds the linear filter that converts time-series data to a series of uncorrelated, random shocks called white noise. Once estimated, the filter may be used to generate optimal extrapolative forecasts, in the sense that all the systematic information present in the history of a time series has been effectively utilized. We refer to HPS, Chapter 4, for a detailed discussion.

4.2. Leads and lags among marketing variables

Few managers would disagree that their marketing actions are effective in more than just the period in which they are taken. This dynamic aspect of marketing is exhibited in two ways: (1) lagged effects, i.e. sales changes, competitive reactions, and other forms of marketing behavior may be noticeable in one or more periods after the original stimulus occurs; and (2) lead effects, i.e. consumer or competitors may anticipate a marketing stimulus and adjust their behavior before the stimulus actually occurs. There has been a great deal of empirical research on lagged effects in marketing and virtually none on lead effects. However, since the methods used in modeling leads and lags are the same, we can discuss them together. (This statement is only correct in a time-series (Box–Jenkins) sense; it is not correct in a traditional econometric sense. In traditional econometrics, lags have a known (and, hence, deterministic) stimulus whereas leads have an unknown (and, hence, stochastic) stimulus.) Furthermore, as the quality of marketing data continues to increase with advances in management information systems, the accurate modeling of marketing dynamics becomes more important. For example, ignoring lagged advertising effects on sales may be more serious on monthly than on annual data if the true advertising duration is a few months.

An important distinction in marketing dynamics is between pulse and step actions in marketing. A pulse action is turned on and off by the marketer, e.g. an advertising campaign or a price promotion. A step action has more of a permanent character, for example, the launching of a new product in the line or the tapping of a different distribution channel. Although dynamic effects are expected to occur in both cases, the literature has for all practical purposes only investigated pulse actions.
Moreover, these studies have generally focused on two marketing-mix elements: advertising and price (in particular, price promotions), although, in principle, all marketing-mix variables can have dynamic effects. We will therefore focus our discussion on advertising and price dynamics. We will also address competitive reactions.

4.2.1. Advertising dynamics

There are various reasons why advertising's effect on sales may be distributed over time. The most important ones are:

- The advertising may not get noticed by the customer until some time after its expenditure. For example, Montgomery & Silk [1972] found lagged journal advertising effects on prescription drug sales of up to six months. This may be caused by physicians delaying the reading of their professional journals.
- An advertising-induced purchase may be followed by subsequent purchases if the product is satisfactory. Likewise, the positive word-of-mouth resulting from the initial purchase may bring new customers into the market. (Gion & Horsky [1990] address the problem of untangling the effects of purchase reinforcement and advertising carryover.)
- In some instances, competitive reaction to an advertising campaign may be slow. If the advertising is effective, it may affect sales performance until competitive retaliation takes place.
- Advertising may gradually build up customer loyalty and thus be responsible for more than the immediately observable short-term sales fluctuations.

The literature on advertising dynamics has addressed two important questions: (1) What are the cumulative advertising effects and (2) does advertising wear out? We address each of these issues in turn.

Cumulative advertising effects. It is difficult to specify the advertising dynamics in a market from marketing or psychological theory alone. Although several approaches are possible, three simple yet intuitively appealing models have been used most frequently. These models all recognize that sales or market-share data are typically autocorrelated, but they differ in opinion on whether or not advertising is causing the autocorrelation in sales, i.e., whether or not an advertising carryover effect exists. Among the most popular models are:

1. The autoregressive current effects model (ACE): This model argues that advertising only has contemporaneous effects on sales:

   \[ Q_t = \beta_0 + \beta_1 A_t + \epsilon_t \]  

   (4.3)

However, other factors such as consumer inertia and habit formation cause sales to fluctuate smoothly over time, which is represented by the autoregressive process of the error term:

   \[ \epsilon_t = \rho \epsilon_{t-1} + \omega_t \]  

   (4.4)
where $\omega_i$ is white noise. The implied advertising carryover effect in the ACE model is zero, so that the short- and long-run impact is the same ($= \beta_1$).

2) The distributed lag model (Koyck): This model arises when advertising has an infinitely long effect on sales, but with an exponentially decaying pattern over time. The short-term effect is $\beta_1$ and subsequent-period effects are $\lambda \beta_1$, $\lambda^2 \beta_1$, $\lambda^3 \beta_1$, ..., $\lambda^n \beta_1$,

$$Q_t = \beta_0 + \beta (1 - \lambda) \sum_{i=0}^{\infty} \lambda^i A_{t-i} + \omega_t$$  \hspace{1cm} (4.5)

where $\lambda$ is the carryover effect of advertising and must be less than 1. This model is estimated by applying the Koyck transformation, to yield

$$Q_t = \beta_0(1 - \lambda) + \lambda Q_{t-1} + \beta_1 A_t + \xi_t$$ \hspace{1cm} (4.6)

where $\xi_t = \omega_t - \lambda \omega_{t-1}$ and $\omega_t$ is white noise. The implied long-term effect of advertising ($N \to \infty$) is $\beta_1/(1 - \lambda)$ and $\theta\%$ of the long-run impact of advertising occurs in $\log(1 - \theta)/\log(\lambda)$ periods [e.g. Russell, 1988].

3) The partial adjustment model (PAM): This response pattern occurs when consumers can only partially adjust to advertising or other marketing stimuli. However, they do gradually adjust to the desired consumption level, which causes the advertising effects to be distributed over time:

$$Q_t = \beta_0(1 - \lambda) + \lambda Q_{t-1} + \beta_1 A_t + \omega_t$$ \hspace{1cm} (4.7)

The partial adjustment model is very similar to the Koyck scheme, except for the structure of the error term. The implied long-term advertising effect is also $\beta_1/(1 - \lambda)$.

Selection among these alternative specifications is done by nesting, a process in which the parameters of lower-order equations are contained within the parameters of higher-order ones. For details on this process, see Bass & Clarke [1972], Weiss & Windal [1980], and especially Leeflang, Mijatovic & Saunders [1992].

The standard linear model assumes that disturbances are independent. The alternative hypothesis may be that two disturbances $s$ periods apart are correlated. The correlation between these disturbances is called the autocorrelation coefficient, $\rho$. In the presence of first-order autocorrelation, a two-step estimation procedure is required. The first step involves obtaining an estimate of $\rho$ by means or ordinary least squares (OLS) estimation. The second step requires that this estimate of $\rho$ be used in an estimated generalized least squares (EGLS) regression. Marketing applications include Simon [1982] and Vanhonacker [1984].

The Koyck model has been the most frequently used among the three most popular models. In an important survey of 70 empirical studies of advertising carryover effects using the Koyck model, Clarke [1976] concluded that:

"the published econometric literature indicates that 90% of the cumulative effects of advertising on sales of mature, frequently purchased low-priced products occurs within 3 to 9 months of the advertisement. The conclusion
that advertising's effect on sales lasts for months rather than years is strongly supported”.

Clarke's conclusion is a very interesting one, as it gets to the heart of a key advertising management question: How long do the economic effects of advertising last? However, when we try to answer the question with econometric techniques, there is a tendency to find different advertising durations for different data intervals, i.e. there may be a data-interval-bias problem. In particular, Clarke finds that researchers using annual data tend to discover multiple-year advertising effects, which is in conflict with his empirical generalization. This report has spurred an active interest in the data-interval-bias problem which is covered in more detail in HPS. The data-interval-bias problem is not a data problem; rather it is a modeling (specification) problem. See Vanhonacker [1991] for an autocorrelation test of the Koyck scheme.

Advertising wearout. More recent attention has focused on asymmetric patterns in the dynamic sales response to advertising. Little [1979]'s five phenomena of advertising response include three dynamic aspects:

1. different rise and decay rates;
2. changing effectiveness over time;
3. hysteresis, i.e. the response may fall off with constant advertising.

Different rise and decay rates of sales response to advertising are also known as the wearout phenomenon. Advertising wearout may occur for two reasons. First, for consumer or industrial durables, there may be a fixed number of potential customers actively looking to buy the product at any point in time. As an advertising campaign is launched, the sales rate increases immediately, but then tapers off because customers leave the market as soon as they purchase the product (i.e. a market-depletion effect). Second, for frequently purchased products, we often observe impulse-response buying, i.e. an immediate reaction to new advertising which disappears even while the advertising is still running (i.e. an adaptation effect). Either way, the response dynamics may be asymmetric, which must be accommodated by a special function.

Simon [1982] proposes a differential-stimulus response model to incorporate the wearout effect. The differential stimulus is the difference between current and previous advertising or zero, whichever is greater:

$$Q_t = \beta_0 + \beta_1 A_t + \beta_2 \max(A_t - A_{t-1}, 0). \tag{4.8}$$

Sales and advertising may be measured in logarithms to incorporate decreasing returns to scale. The wearout hypothesis is tested by a positive coefficient for $\beta_2$. This implies that, whenever an advertising campaign is new, there will be an extra response effect above and beyond the level stimulus effect ($\beta_1$). Simon tested the model successfully on three frequently purchased products. Also, Hanssens & Levien [1983] found differential stimulus effects in print and television advertising for the US Navy manpower recruitment program.
The managerial implications of advertising wearout are interesting. Simon [1982] argues that the advertising budget should be allocated to pulsing and constant-spending budgets and that the share of pulsing increases with the differential-stimulus effect. The most profitable advertising strategy is one of 'alternating pulsation', i.e. a pulse in every other period. We do not know, though, what the best length of a pulsing period is.

In conclusion, several empirical studies have demonstrated the existence of dynamic effects of advertising on sales performance. The lag lengths of the effects are several months, although the generalization can only be made for mature, low-priced, frequently purchased products [Clarke, 1976]. Furthermore, advertising wearout has been observed in several cases. Both phenomena have important managerial implications for the optimal timing of advertising efforts.

4.2.2. Price dynamics

Price-setting is complex, involving cost, demand, competitive, and organizational considerations. In many instances the resulting price is stable over time, perhaps indicating that an equilibrium has been reached. In other cases there is frequent use of temporary price changes such as price promotions or seasonal price hikes. Empirical research on price dynamics has focused mostly on the latter.

The simplest form of consumer response to a price change is the zero-order model, i.e. sales are a function of the current price only, which implies a static response function. Zero-order price response has been observed on several occasions, even when dynamic price effects were specifically tested. For example, price response for eleven brands of an Australian household product was found to be strictly zero-order [Carpenter, Cooper, Hanssens & Midgley, 1988].

Zero-order price response implies that there is no transient component in the sales–price relation. Buyers are reacting to current prices only; they are not taking advantage of temporary price-cuts by stocking up on the product, nor are they anticipating future price movements. Thus, competitors are operating along a static demand curve and can only influence the level, not the timing of their customer's purchases.

If consumers deviate at all from this myopic behavior, dynamic price response should occur. The most common form is stockpiling, i.e. moving future purchases to the present to take advantage of a temporary price-cut. This leads to an asymmetric response function not unlike Simon's differential-stimulus model.

A more sophisticated dynamic consumer response is to anticipate a future price-cut and thus to reduce purchasing levels until the price reduction occurs. This scenario is described by Doyle & Saunders [1985] as 'regret reduction'. Their empirical example on a European supplier of natural gas and gas appliances did reveal such anticipations, not at the customer level, but rather at the sales-force level. Salespeople were taking advantage of planned promotion campaigns by enticing customers to switch their purchases into the promotion period so that they would receive higher commission rates.

Dynamic price response was studied more formally by Winer [1985], using a vector-price model which distinguishes between five price concepts: anticipated
price, price uncertainty, unanticipated inflation or deflation, and reservation price. In an empirical test of seven consumer durables, several of the price concepts were statistically significantly related to probability of purchase. Similarly, a price expectations effect was found to exist in an analysis of coffee buying [Winer, 1986]. These and other studies highlight the importance of using dynamic models for assessing consumers' price-responsiveness.

Price dynamics have also been studied at the firm level. In particular, DeSarbo, Rao, Steckel, Wind & Colombo [1987] developed a friction model for describing and predicting price changes. The model posits, first, that in the absence of major frictions, companies will tend to hold their price constant. However, upward tensions such as inflation and downward tensions such as competition may build up in price-setting to a point where a threshold is exceeded, which prompts the firm to adjust its price (upward or downward). The resulting threshold model of price movements over time is estimated using maximum likelihood. An application to weekly mortgage interest-rate setting revealed that an individual bank will adjust its interest rate in response to the previous weeks' changes in the cost of money and competitive rates.

4.2.3. Reaction functions

Construction of the sales-response function is just one step in marketing programming. If competitors react to our actions, we must develop competitive reaction functions. In a path-breaking article, Bass [1969] presented a model of a market mechanism for cigarettes that contained sales-response functions and the corresponding advertising decision rules for the filter and non-filter segments. The advertising decision-rule equations indicated how one segment's sales influenced the other segment's advertising expenditure decision in the same year. This study was followed by a major investigation of the ready-to-eat cereal market [Bass & Parsons, 1969]. Again there was a four-equation model of the market mechanism. This time there was a pair of relationships describing the behavior of a particular brand and another pair for all other brands combined. A major empirical result was that competitive advertising activities seemed to stimulate primary demand in this market. The advertising decision rules were based on marketing actions and results in the previous bimonthly period. Samuels [1970/71] tested a model very similar to that of Bass & Parsons with data from three markets: scouring powders, toilet soaps and household cleansers. Wildt [1974] studied retail price advertising as well as network advertising as a managerial decision rule. Since three firms accounted for the majority of sales in the industry, his market mechanism contained nine equations.

Lambin [1970] and Lambin, Naert & Bultez [1975] investigated a small electrical appliance from the Stackelberg leader–follower perspective. They constructed the followers' reaction functions for advertising, price and quality. They found that while the competition did react to a change in one marketing instrument by the leader with a change in the same instrument, the competition also changed other elements in their marketing mix. For instance, the competition might only partially match a price-cut and, instead, increase their advertising outlays. In the
discussion of their results, Lambin, Naert & Bultez raised the possibility of kinked demand curves coming about if the competition did not react when it would result in increased market share. However, they did not address this in their ordinary least squares estimation of the reaction functions.

Parsons & Schultz [1976] describe the general form of these models, which they call models with endogenous competition. Our interest is in a typical decision rule for a firm. The level of a particular decision variable for a specific competitor may be affected by the firm’s own managerial decision variables, by the managerial decision variables of each of its competitors, by its own and its competitors’ past sales, and by autonomous environmental conditions including seasonality. What is important to recognize is that the reaction function of the economist can be, and often is, embedded in the more general construct of the decision rule.

Hanssens [1980b] incorporated the phenomenon of the level of one marketing instrument affecting, or being affected by, levels of other marketing instruments within the same firm into a generalized reaction matrix. This matrix is partitioned so that the main diagonal blocks represent simple competitive reaction. The off-diagonal blocks represent multiple reaction. The diagonal elements represent intra-firm effects and the off-diagonal ones represent inter-firm effects. Hanssens noted that in an oligopoly the inter-firm reaction elasticities should be zero if the firm is a follower (Cournot–Bertrand reaction function) and nonzero if the firm is a leader (Stackelberg reaction function).

Most of what little evidence we have on reaction elasticities is due to Lambin [1976], Metwally [1978] and Leeflang & Wittink [1992]. Leeflang & Wittink conclude:

“Our findings show that simple competitive reaction functions would fail to capture systematic effects due to marketing variables. At the same time, simple reactions do account for a disproportionate number of reaction effects. Importantly, the estimated competitive reactions appear to be very complex”.

Whether a recursive or nonrecursive model is necessary to represent the system containing both sales-response functions and reaction functions depends largely on the data interval. As the data interval lengthens, more mix movements would appear to be simultaneous. Temporal interrelationships can be identified using time-series analysis [Hanssens, 1980a; Leeflang & Wittink, 1992]. The functional form of reaction functions examined has been restricted to either linear or multiplicative. That firms react to either an absolute change or a relative change in competitive behavior seems reasonable; however, no attempt has been made to show which of these is true.

4.3. Assessing the direction of causality

Determining the direction of causality may be straightforward when only two variables are involved, but real-world marketing systems are often so complex that the causal chains cannot be easily established a priori. For example, in competitive markets causal relations may exist in many directions among the following
variables: product sales, industry sales, market share, profits, marketing efforts, competitive marketing efforts, and environmental conditions. Although in this case we would have a good idea of the elements in the information set (level 0), it would be difficult to posit one structural marketing model from prior insight alone.

This section discusses ETS (econometric and time-series analysis) model-building techniques for the data-driven assessment of causality in marketing systems. We introduce the concept of Granger causality, discuss empirical testing procedures, and summarize the use of these techniques in marketing models to date.

4.3.1. The concept of Granger causality

It is difficult to establish a workable definition of causality in nonexperimental research. As far as statistical analysis is concerned, we often hear the remark that 'correlation does not imply causation'. But when we adopt a stochastic view of time-series behavior, temporal ordering of events can be used to make an empirical distinction between leading and lagging variables. That distinction is at the basis of a well-known definition of causality due to Granger [1969].

Suppose a marketing system is defined by a two-variable information set \((X, Y)\).

In an attempt to forecast \(Y\), we could build a univariate model, i.e., considering the past of \(Y\) alone, or we could combine the past of \(Y\) and the past of \(X\) in a bivariate model. Now, \(X\) is said to Granger cause \(Y\) if the mean squared forecast error (MSFE) of \(Y\) using the bivariate model is smaller than the MSFE of the univariate model. Formally:

For the information set containing \(X\) and \(Y\), \(X\) is said to Granger cause \(Y\) if:

\[
\text{MSFE}(Y_t | Y_{t-1}, \ldots, Y_{t-k}, X_{t-1}, \ldots, X_{t-m}) < \text{MSFE}(Y_t | Y_{t-1}, \ldots, Y_{t-k}),
\]

where \(k\) and \(m\) are positive integers indicating the maximum memory length in \(Y\) and \(X\).

There are three distinctive components to Granger's definition:

- It stresses the importance of an adequately formulated information set.
- The empirical detection of causality between \(X\) and \(Y\) is valid only insofar as no major factors \(Z\) are missing from the information set. The 'null' model against which forecasting performance is evaluated is a powerful rival. For example, univariate time-series models have been shown to outperform complex econometric models of the US economy [Nelson, 1972].
- The ultimate test is done out-of-sample. Thus, statistical significance of transfer-function parameters alone is not sufficient to establish Granger causality.

Granger causality applies well in a marketing context. For example, monthly time series of the number of airline passengers on a route have often been found to follow an ARIMA \((0,1,1)(0,1,1)_{12}\) process – known as the airline model – which predicts future passenger levels remarkably well. This model may be written as

\[
\begin{align*}
\hat{z}_t &= z_{t-1} + z_{t-12} - z_{t-13} - \theta_1[z_{t-1} - \hat{z}_{t-1}] \\
&\quad - \theta_{12}[z_{t-12} - \hat{z}_{t-12}] + \theta_{13}[z_{t-13} + \hat{z}_{t-13}]
\end{align*}
\]
where $z_t$ is passengers at time $t$, $\hat{z}_t$ is time $t$ passengers predicted at time $t - 1$, and $\theta_i$ are forecasting parameters ($i = 1, 12, 13$). The marketing question 'does manipulating the air fares affect demand?' might be poorly answered by merely correlating or regressing passenger and air-fare series. Granger’s definition would assess whether or not air-fare information improves the prediction of passenger levels beyond what is achieved by extrapolation. If the airline pricing managers act rationally and forecast demand accurately, the air fares may follow a rigid seasonal pattern with little extra variation. In that case we may well find that they do not Granger cause passenger demand, but, instead, are caused by (perfectly anticipated) passenger movements. One extension of the definition includes present as well as past values of $X$ in the prediction of $Y$. This is known as ‘Granger instantaneous causality’ and is more difficult to measure empirically [Layton, 1984].

4.3.2. Test procedures

Although the concept of Granger causality was developed in economics, it did not achieve recognition until time-series methods for its execution became available. Several procedures have been proposed, including the double prewhitening technique and two regression-based methods due to Granger & Sims.

The double prewhitening method, first proposed by Haugh [1976] and later extended by Haugh & Box [1977], Pierce [1977] and Pierce & Haugh [1977], establishes the direction of causality between two series by cross-correlating the residuals of univariate time-series models fitted to each. How does this method relate to the definition of Granger causality? In the first stage, the predictive power of each series’ past is removed via the prewhitening operation. Then, by cross-correlating the residuals at various lags, the method scans the data for any additional sources of covariation. If a significant cross-correlation exists at positive or negative lags, it contributes to Granger causality in that direction. If the spike occurs at lag 0, it contributes to Granger instantaneous causality, but the direction of the effect cannot be established by itself. The main restriction of double prewhitening, though, lies in the fact that both stages are typically carried out on the same sample, so there is no true forecasting test. That limitation prompted Ashley, Granger & Schmalensee [1980] to develop a supplementary test for the out-of-sample performance of univariate vs. bivariate time-series models.

The double prewhitening method has been instrumental in stirring controversial debates of cause and effect in the macroeconomic and financial economics literature. For example, Pierce [1977] reported a lack of relations among several key interest and money indicators previously thought of as highly interrelated. In marketing it was first used to establish primary demand vs. market-share effects of airline flight-scheduling and advertising and to sort out various patterns of competitive reactions among airlines [Hanssens, 1977, 1980b]. Causality testing has also been extensively discussed in Bultez, Leeflang & Wittink [1991]. A more comprehensive overview of causality tests in marketing may be found in HPS, Chapter 5.

Regression models have been used as well, in particular techniques attributed
to Granger [1969] and Sims [1972]. By regressing current Y against lagged Y and lagged X [Granger], or against past and future X [Sims], we may test for causality without the preliminary step of univariate time-series analysis. Regression-based causality tests were first used in marketing to determine the causal ordering of advertising and aggregate consumption. Using the Sims method, Jacobson & Nicosia [1981] established a contemporaneous relation between annual advertising and personal consumption in the UK. They also found that advertising affected next year's consumption, and vice versa. Their regression results were confirmed by a double-whitening test on the same data.

4.4. Shape of the sales response function revisited

While we have already introduced some dynamic properties of sales-response functions such as asymmetry in response and coefficient variation, we now wish to discuss them in more detail. Asymmetry in response occurs when the magnitude of response to a change in a marketing instrument is different depending on whether the change is an increase or a decrease. This is different from asymmetry in competitive effects, that is, a change in a brand's marketing effort affecting each competitive brand differentially. Coefficient variation occurs when a coefficient changes over time, e.g., an advertising elasticity over the product life-cycle.

4.4.1. Asymmetry in response

The magnitude of sales response to a change in a marketing instrument might be different, depending on whether the change is upward or downward. The effect is beyond any that might be explained by the nonlinearity of the sales-response function. Sales might rise quickly under increased advertising, but stay the same or decline slowly when advertising is removed. As noted previously, this phenomenon has been termed 'hysteresis'. One explanation is that higher advertising expenditures create more customers through greater reach as well as greater frequency. Under the customer holdover paradigm, these new customers subsequently repurchase the product. Thus, if advertising is cut back, sales will fall by less than would be the case in the absence of this effect. (Sasieni [1989] observes that 'In 20 years of studying response data [he] can recall only one example of hysteresis and that concerned price rather than advertising.') Sales response to price may also be asymmetric because of habit formation. The sales-response function will be kinked. A price rise would be less elastic than a price fall. This is termed 'addiction asymmetry'.

Asymmetry in response is usually captured by ratchet models. There are two types of ratchet models. The first is saw-toothed in appearance. The sales-response function is kinked at the prevailing level of a marketing instrument, irrespective of past changes in the instrument. Segments of the adjustment path for increases in level are parallel; segments for decreases are also parallel to each other. The second resembles a bird's footprint in appearance. Purchasing habits that were developed by the use of a product under a condition of record-breaking marketing activity will not be broken easily if marketing effort recedes. Rachet or rachet-type
models have been used by Parsons [1976] for advertising carryover effects, by Young [1983] in a cigarette sales-price investigation, and by Simon [1982] in his ADPULS model of the advertising wearout phenomenon [see (4.8)]. A continuous-time version of ADPULS is discussed by Luhmer, Steindl, Feichtinger, Hartl & Sorger [1988a, 1988b]. Asymmetric models may be considered a special case of coefficient variation.

Asymmetry can arise in frequently purchased branded goods because of the phenomenon of fast learning and slow forgetting on the part of consumers. Asymmetry in response to advertising has been addressed by Parsons [1976], Haley [1978], Little [1979] and Simon [1982]. Haley reported on some experiments that showed an immediate sales response to increased advertising. In addition, these experiments indicated that even though the advertising was maintained at the new and higher levels, the magnitude of response gradually became less and less. Little offered two explanations for this. One is that advertising causes prospects to try a product. Only a portion of these new triers became regular purchasers. Consequently, sales taper off from their initial gain to a lower level. (It might be argued that this is not a response phenomenon at all. Rather it might be considered a sampling problem, or more generally, an aggregation problem, because time-series observations are derived from different populations.) The second explanation is that the advertising copy is wearing out. We believe another possible explanation would be competitive reaction.

Asymmetry in response to price has been discussed by Moran [1978]. He provides a summary of some price research that has been conducted in a variety of consumer-product categories. He argues that the only way to analyze a price elasticity is in terms of relative price. Relative price expresses a brand's price relative to the average price for the product category in which it competes. One of his major findings is that a brand's upside demand elasticity and downside elasticity can differ. He conjectures that one reason these elasticities might differ is that consumer segments are not equally informed about what is going on. For instance, an unadvertised price change is more likely to be noticed by current customers. We must note that Moran was working with data that are obsolete now that scanner data are available, and that price promotion and price effects are different.

Russell [1992] provides a generalization of much of the work on brand price competition. His model, the Latent Symmetric Elasticity Structure (LSES), assumes that the market-share cross-price-elasticity $\eta_{ij}$, the percentage change in the share of brand $i$ with respect to a 1% change in the price of brand $j$, is equal to the product of the (asymmetric) clout factor of brand $j$ and a symmetric index of the substitutability of the brand pair $(i,j)$. His empirical work shows that clout factors depend upon both market share and average price while the pattern of substitution indices is influenced by the brand's average price level. Assuming that price is correlated with quality, this work suggests that both the pattern of asymmetry (explained by the clout factors) and the draw pattern from price promotions (explained by substitution indices) depend on quality levels. Also see Bucklin, Russell & Srinivasan [1992].
4.4.2. Time-varying coefficients

The effectiveness of each controllable marketing instrument is frequently treated as having the same value over time. However, market response to managerial decision variables might change because of the impact of marketing actions by either a company or its competitors or because of shifts in the environment. For example, Simon & Sebastian [1987] allow the 'innovation' or 'imitation' coefficients in the Bass market-growth model to vary systematically with advertising.

If structural changes occur at known points in time, the changes in the coefficients of the relevant variables can be represented by dummy variables. A dummy variable takes the value 1 when a phenomenon or characteristic is present and the value 0 when it is absent. Palda [1964] assumes that restrictions placed upon Lydia Pinkham's advertising copy by the Food and Drug Administration in 1914 and again in 1925 and by the Federal Trade Commission in 1940 could be captured by dummy variables. These dummy variables affect only the intercept of the sales response function. A somewhat more appropriate approach might have been to use the dummy variables to model changes in the slope coefficient, i.e. the effectiveness of advertising.

Unfortunately the timing of a structural change is rarely known. The parameters in a response function might vary between each time period rather than only between a few time periods. If the parameters follow some random process, the parameter variation is said to be stochastic. Any model of stochastic parameter variation that assumes an autoregressive process is called a sequentially varying parameter model. If the parameters themselves are functions of observable variables, the parameter variation is said to be systematic. Coefficients can also vary over cross-section units.

Specific models, e.g. switching models, have been developed for situations where sample observations are generated by two or more distinct regimes. A switch from one regime to another may depend on time, but alternatively, it might depend on a threshold value for some variable or occur stochastically. In the two-regime case, the switching model can be written as

\begin{equation}
\text{Regime 1: } Q = f(X_1; \beta) \text{ if condition holds,}
\end{equation}

\begin{equation}
\text{Regime 2: } Q = f(X_2; \gamma) \text{ if condition does not hold.}
\end{equation}

Bemmaor [1984] tests for the existence of an advertising threshold effect using the stochastic version of this model. The first-regime condition is 'with probability \( \theta \)' while the second-regime condition is 'with probability \( 1 - \theta \)'. Lee & Brown [1985] study the impact for Florida Department of Citrus coupon promotional programs on the demand for frozen concentrated orange juice. Separate sales-response functions were estimated for coupon users and nonusers. The probability of a household redeeming a coupon was itself a function of household characteristics, market conditions, and properties of various promotional programs. The modeling of changing market environments is critically reviewed by Wildt & Winer [1983].

One source of dynamic systematic variation in marketing is the product life-cycle. Marketing theory states that the demand elasticities of managerial decision
variables change over the product life-cycle. The theory has been interpreted to say that the advertising elasticity is highest at the growth stage of a product life due to the need to create increased product awareness, and lowest during maturity, with elasticities increasing slightly through saturation and decline stages of the product life-cycle [Mahajan, Bretschneider & Bradford, 1980]. The theory supposedly conjectures that the price elasticity increases over the first three stages – introduction, growth and maturity – and decreases during the decline stage [Mickwitz, 1959]. Very early in the product life-cycle the demand curve may be relatively inelastic as consumers purchase a product to learn about its unknown quantities [Tonks, 1986]. Over time as information accumulates consumers become more price-conscious. General economic principles predict that the presence of close substitutes causes high elasticities. Thus, as a product matures, price elasticities should increase (become more negative) because of the availability of close substitutes. Toward the very end of the product life-cycle, the demand curve may (again) become inelastic as only 'diehard' brand-loyal customers remain.

Empirical evidence on changes in the efficiency of various marketing instruments at different stages of the product life-cycle is sparse. Indications are that advertising elasticities generally fall as products pass through their life-cycles [Parsons, 1975; Arora, 1979]. A naive model of this process would be

\[ e_d = \alpha_1 e^{-\alpha t} + \alpha_3. \]  

(4.12)

Using 'moving-window' regressions, Leeflang, Alsem & Reuyl [1991] investigate the competition among suppliers of advertising media in the Dutch advertising market. They find that the cross-elasticity of television advertising on newspaper advertising declined over a 20-year period ending up not significantly different from zero.

Simon [1979] in an empirical study of 35 brands in seven different markets found price elasticities seem to decrease markedly during the introduction and growth stages, reaching a minimum in the maturity stage, after which they may experience an increase during the decline stage. For industrial chemicals, Lilien & Yoon [1988, p. 273] conclude:

"The level of price elasticity tends to be lower during the later stages of the product life cycle (maturity and decline) than during the earlier stages (introduction and growth). There is no clear tendency of shift in the level of the price elasticity between the introduction and growth stages. Over the latter two stages of the product life cycle (maturity and decline), price elasticity shows a tendency to be stable".

Although these results may be tentative due to methodological problems, e.g. Shoemaker's [1986] comment on Simon [1976], nonetheless, these empirical findings do seem inconsistent with current marketing theory. Snell & Tonks [1988] provide more conventional results in their examination of eight products in the UK chocolate confectionery industry. They report that, with only a few exceptions, elasticities become larger (more negative) over the product life-cycle.
5. Empirical findings and marketing generalizations

A marketing manager's central concern is how selective marketing activities of a brand affect its sales. The manager would like to draw on accumulated wisdom about the shape of the response function. The manager also recognizes that sales effects come through changes in selective demand, through changes in primary demand, or both. As a consequence, the manager might well want to couple a model of industry demand with a market-share model. Whether interested in the brand level or the industry level, the manager would like to know what marketing generalizations have been discovered. This is especially helpful for a manager confronting a new market/product situation.

5.1. Empirical evidence on the shape of the response function

The shape of response to a particular nonprice marketing instrument, with the remainder of the marketing mix held constant, is generally concave. Sales always increase with increases in marketing effort, but exhibit diminishing returns to scale. Sometimes the sales response function might be S-shaped with effort. Initially sales may exhibit increasing returns to scale and then diminishing returns to higher levels of marketing effort. Jones [1984] emphasizes that 'nowhere is it suggested that increasing returns are anything but limited and temporary'. The temporary increasing-return phenomenon may be localized in the introductory stage of the product life-cycle and related to increasing distribution coverage, i.e. an improvement in the conversion of demand into sales [cf. Steiner, 1987]. The shape of the response function might be different depending on whether marketing effort is increasing or decreasing, that is, response may be asymmetric. The shape of the response function also might vary with changes in the environment.

The preponderance of empirical evidence favors the strictly concave sales response to nonprice marketing-decision variables. This is especially true for mass-media advertising of frequently purchased goods. For instance, Lambin [1976, p. 95], after doing an analysis of 107 individual brands from 16 product classes and 8 different countries of Western Europe, concludes that 'the shape of the advertising response curve is concave downward, i.e. that there is no S-curve and no increasing returns in advertising a given brand by a given firm'. Earlier, Simon [1970, pp. 8–22] had surveyed the evidence then available on the shape of the response function and found that 'both sales and psychological [nonsales measures of behavior] suggest that the shape of the advertising-response function is invariably concave downward, i.e. that there is no S-curve'. Reviews by Simon & Arndt [1980] and Aaker & Carman [1982] also indicate diminishing returns to advertising.

There are several reasons to expect diminishing returns to increased advertising expenditures [Jagpal, Sudit & Vinod, 1979]. For one, the fraction of unreached prospects is progressively reduced as advertising increases. Consequently, most of the impact of additional advertising messages at high levels of advertising takes place by means of increased frequency. Moreover, after a small number of exposures, perhaps as few as three, increased frequency has very limited marginal
effectiveness. Grass & Wallace [1969] among others report on the satiation effects of television commercials. Ottosen [1981] proposed a theory of the individual's purchase-response function, and on the basis of this theory, he concludes: 'as advertising effort is being increased, returns in sales must generally be expected to diminish'.

An S-shaped sales response to advertising has long been conjectured [Zentler & Ryde, 1956]. However, this proposition has not been tested explicitly. Two studies explore the proposition that the relation between market share and advertising is S-shaped. Johansson [1973] found for a women's hairspray that the advertising effect was concave rather than the proposed S-shape. Rao & Miller [1975] adopt an ad hoc procedure to develop S-shaped response functions for five Lever brands. The work of Ambar Rao & Miller seems suspect, however, since they discard markets that were 'out of line'. This means that for the two brands they discuss in detail, 27% and 20% respectively, of the markets were omitted. Eastlack & Rao [1986] also apply Rao & Miller's methodology. A linear sales response to radio gross-ratings points and television gross-ratings points was estimated for each Selling Areas Marketing, Inc. (SAMI) market. Gross-ratings points (GRP's) refers to the total number of exposures generated by an advertising schedule. It is a measure of delivered advertising. Inspection of per-capita estimates of marginal response to radio GRP levels revealed no significant response below 180 GRP's (an indication of a threshold), a sharp increase in response as GRP's increased between 180 and 230, a slight (almost flat) decline in response as GRP's increased from 230 to 340 (an indication of saturation), and low response to a few observations with GRP's above 400 (an indication of supersaturation?). These two works, unfortunately (because of their ad hoc nature), are the only published support for an S-shaped response function. The possibility of an S-shaped relation between market share and communications (advertising plus personal selling) for an industrial product is considered by Gopalakrisna & Chatterjee [1992]. Their empirical evidence implies a concave response function [p. 191, fn. 3]. Broadbent [1984, p. 310] in a discussion of an advertising stock model reports that "The uncertainty in the data also makes it difficult – we would say from our experience impossible – to prove or disprove the reality of an S-shaped or step function."

The lack of evidence for an S-shaped curve has an important implication for the timing of advertising expenditures. An advertiser might want to choose between two alternative policies, a constant spending rate per period or a pulsed expenditure. Ambar Rao [1970, p. 55] defines a pulsing policy as a pattern of advertising where periods with high advertising intensity alternate with very little or no advertising. A sufficient condition [Rao, 1970, p. 5] for adopting a pulsing policy would be that the sales-response function be S-shaped and the budget constraint be binding. The budget constraint has to require that the alternative constant-rate policy be in the region of increasing returns to scale. But most empirical evidence says that a typical brand has a concave sales-response function; consequently, the S-shape cannot be used to justify a pulsing policy.
The relationship between market share and share of retail outlets seems to be S-shaped. Cardwell [1968] reports that in marketing gasoline incremental new outlets were substantially below average in gallonage until a certain share of market was achieved. Above this critical market-share, performance improved markedly. Lilien & Rao [1976] also postulate an S-shaped relationship between share of market and share of outlets. Neither study provides empirical evidence supporting its claims. Naert & Bultez [1975] do an analysis of the effect of market share on the distribution network of a major brand of gasoline in Italy. Their results support the S-shaped hypothesis at the market-share level. However, when the hypothesis is tested at the aggregate brand-switching level, it is rejected. In any event, the relationship between market share and share of outlets may be simply an expression of the difference between demand and sales. In general, an S-shape for market share may be an artifact of its being constrained to lie between zero and one. The underlying sales-response function may not be S-shaped.

If support for S-shaped sales response is weak, even less support exists for the threshold effect. Although many marketing managers believe that a threshold effect operates within their market [Corkindale & Newall, 1978], Simon [1970, p. 22] expresses the opinion that “threshold effects ... constitute a monstrous myth”. Even though the argument might be made at the individual level that a prospect might be unaware of a brand or unwilling to buy it until several advertising messages have been received, little evidence of this threshold phenomenon in aggregate sales-response functions has been found. Corkindale & Newall [p. 373] note that ‘Little generalisable evidence of either phenomena [threshold and wearout levels of expenditure] seems to exist. This is mostly because managers and their agencies avoid operating at or near the supposed limits.’

A most interesting attempt to identify a threshold effect in an aggregate sales response is by Bemmaro [1984]. A market-share response function is partitioned into two regimes – above and below the threshold [see (4.1)]. A multiplicative function describes each segment. A random shift between these two regimes was postulated. For the product studied, the best fit occurs when the estimate of the proportion of observations above the threshold was 73%. The corresponding threshold advertising share was deduced to be about 18%. Thus, these results indicate decreasing returns to scale but with a discontinuity.

The existence of a saturation level is universally accepted. Nonetheless, the saturation level is rarely explicitly modeled and measured. The usual procedure is to represent response by a function that allows any given level to be surpassed, but requires increasing effort to exceed each higher level. This approach is probably adequate for use in decision models focusing on short-run marketing tactics; however, when interest is in long-term strategy, the saturation ceiling should be estimated. One industry sales-response function in which the saturation level was explicitly modeled was Ward's [1975] study of canned, single-strength grapefruit juice. He used a reciprocal model. Saturation sales, \( Q^0 \), were estimated to be 69.82 million gallons. The highest sales observed to date were 53.77 million gallons.

The notion of a supersaturation effect, excessive marketing effort causing reduced
sales, has been promulgated by Ackoff and his colleagues [Waid, Clark & Ackoff, 1956; Rao, 1970; Ackoff & Emshoff, 1975] and is being incorporated into marketing theory [Enis & Mokwa, 1979]. A possible explanation could involve the content of the advertising message. If the message is inappropriate, more exposure to the 'wrong' message could have a negative effect. This could also be the case with the 'wrong' advertising-copy execution. In particular, some advertisements may be very irritating after too many exposures. Still, the argument for supersaturation in advertising is unconvincing. Campaigns such as Wisk's 'Ring around the collar' have been very much disliked by viewers, but, nonetheless, have been very successful in terms of sales. The only empirical evidence even tangentially bearing on the existence of such an effect comes from Ackoff's Budweiser study. While previous research, such as that of Parsons & Bass [1971], has shown that reducing advertising expenditures may increase profits even though sales are lost, the Budweiser study is the only research in which reducing advertising not only increased profits but also increased sales. Haley [1978] did report on another beer test in which those areas where advertising was stopped showed better results than the remaining areas. However, subsequent investigation revealed that local distributors, upon finding their advertising support dropped, invested their own funds in advertising. Their efforts more than offset the cuts made by the manufacturer. Participants in the Budweiser study have asserted that adequate controls were maintained in their work; consequently, their results remain an anomaly. Even if supersaturation does exist, it is well outside the usual operating ranges for marketing instruments since management has little incentive to operate even at saturation. Of course, a firm could operate in this region by mistake.

A more plausible argument for supersaturation might be made for marketing-decision variables other than advertising. For example, a reduction in sales-force size might lead to more effort, and consequently sales, if territories and hence potential were realigned and salespeople were more highly motivated because of this. Also, it was recently reported in the business press that a major computer manufacturer increased its sales by decreasing the number of retail outlets carrying its personal computers. The explanation was that with fewer dealers there was less price competition, higher retail prices and hence increased margins, and thus more funds available to each retailer to support direct sales effort. These examples still may not justify the theoretical existence of supersaturation. They may only demonstrate that it could be empirically found if the model is misspecified by ignoring sales-force motivation and/or territory alignment. Whatever the case, there have been no empirical studies of supersaturation for sales-force or distribution variables.

One source of systematic parameter variation is the interaction of the marketing decision variables with each other. Advertising expenditures often influence the magnitude of price elasticity [Moran, 1978; Sunoo & Lin, 1978]. Conventional wisdom is that advertising decreases price-sensitivity. Schultz & Vanhonacker [1978] provide some empirical support for this proposition, yet Wittink [1977a] gives some evidence that relative price becomes more elastic as advertising share increases. The implication is that advertising tends to increase the price-competitiveness of the brand investigated. This supports earlier findings [Eskin,
1975; Eskin & Baron, 1977] that a high advertising effort yields a higher price elasticity than a low advertising effort. Farris & Albion [1980] suggest that the concept of vertical market structures might reconcile what appears to be conflicting evidence. They posit that the relationship between advertising and price depends on whether price is measured at the factory level or at the consumer level. Krishnamurthi & Raj [1988] found that, for a well-established brand, increased noninformational advertising of the mood type decreased price sensitivity. Popkowski, Peter & Rao [1990] postulated a model in which local advertising makes demand more price sensitive while national advertising makes demand less price-sensitive. They reported empirical evidence supporting their model.

Moreover, many secondary dimensions of marketing variables are only operative when the primary dimension of the variable is present. If no advertising expenditure is made, the advertising copy can have no impact on sales. Samples and handouts are distributed in conjunction with a sales call. Parsons & Vanden Abeele [1981] demonstrated that the effectiveness of the calls made by the sales force of a pharmaceutical manufacturer for an established ethical drug varied systematically as a function of collateral material, such as samples, used.

Systematic variation might occur over individuals or territories. Moran [1978] found that the price elasticities for a brand varied from market to market and from segment to segment. Wittink [1977b] tested one brand to evaluate whether demographic variables explained differences in the estimated parameters of the sales response functions for various territories. He found that they did not. Elrod & Winer [1979] had only somewhat better luck in relating household characteristics to the estimated parameters in purchasing response functions for different households. Gatignon & Hanssens [1987] reported that marketing effectiveness in Navy recruiting is inversely related to environmental conditions, in particular, civilian employment rate.

5.2. Marketing generalizations

Guidance for marketing action comes from regularities in aggregate response behavior across markets. Of course, marketing is very product-market-specific and these generalizations are not too precise. We begin with brand-level generalizations, then turn to industry-level generalizations.

5.2.1. Brand-level generalizations

Many brand-level generalizations relate to short-term elasticities. One of the first marketing generalizations [Leone & Schultz, 1980] is that the elasticity of selective advertising on own-brand sales is positive but low. This is supported by the meta-analyses conducted by Aaker & Carman [1982] and Assmus, Farley & Lehmann [1984]. The value for the advertising elasticity with unit sales appears to be of the order of 0.10.

There is a growing body of information on price elasticities. The elasticity of price on own-brand sales is negative and elastic. A meta-analysis conducted by Tellis [1988] found a mean own-price elasticity of about \(-2.5\). (The magnitude
of cross-price elasticities is roughly 0.5, e.g., Bolton [1989].) Some argue that Tellis’s own-price elasticities appear too low. If a simple constant-elasticity model is used, a $-2.5$ elasticity would imply that grocery markups should be 67%. One never sees these types of markups. The reason is that the retailer worries about lost customers, which is not analyzed by Tellis. The large discrepancy in the magnitudes between these two own elasticities – advertising and price – has led to a debate about whether or not price is a superior tactic to advertising [Broadbent, 1989; Tellis, 1989; Sethuraman & Tellis, 1991].

Blattberg & Neslin [1989] propose some generalizations about sales promotions. One is that sales promotions have a dramatic immediate impact on brand sales. This proposition is supported by the work of Chevalier [1975], Gupta [1988] and Moriarty [1985, p. 42]. A second generalization is that short-run promotional cross-elasticities are asymmetric. They mean that the effect of Brand A’s promotion on Brand B’s sales is not the same as the effect of Brand B’s promotion on Brand A’s sales. Blattberg & Wisniewski [1989] found that higher-price, higher-quality brands steal share from brands in the tier below as well as from other brands in the same price-quality tier. However, lower-tier brands do not steal a significant share from the tiers above. A third generalization is that each promotional tool has its own impact and may, in addition, interact with other promotional tools. Blattberg & Neslin point out that promotions are often composed of two or more promotional tools; e.g., price-cut, feature and display may be used together. It may, however, be difficult to estimate the separate effects of each tool because of the high correlations among these variables, i.e., multicollinearity. Blattberg & Neslin [1989] and Guadagni & Little [1983] provide support for this proposition. The last generalization is that promotional price-cut elasticities are larger than those for regular-price elasticities.

Another marketing generalization identified by Schultz & Leone is that increasing store-shelf (display) space has a positive impact on sales of nonstaple grocery items. They found support for this statement in the works of Pauli & Hoecker [1952], Mueller, Kline & Trout [1953], Progressive Grocer [1963, 1964], Cox [1964, 1970], Kotzian & Evanson [1969], Frank & Massy [1970], Kennedy [1970] and Curhan [1972, 1974a, 1974b]. Criticisms of these studies have been made by Peterson & Cagley [1973] and Lynch [1974]. They raise the possibility that the relationship between sales and shelf space should be expressed in terms of a simultaneous system of (nonlinear) equations.

Another possible marketing generalization might be that personal selling has a direct and positive influence on sales. Lambert [1968] found that sales volume of medical X-ray film in a district was related to the number of salespeople employed by the company in the district as well as to a product-mix measure and a selling-price index. However, the direction of causality is not clear. Waid, Clark & Ackoff [1956] in their analysis of the lamp division of General Electric indicated that the number of calls was the only variable to influence dollar sales. Turner [1971] refined the concept of calling effort by defining it as the product of the number of calls and the number of people seen per call. Calling effort was shown to have a significant impact on the actual sales to individual customers. Beswick & Cravens
[1977] reported that dollar sales of one firm's high-priced consumer goods were determined by the salesperson's percentage of time spent in a geographic area and variables representing potential, workload, company experience, salesperson experience, and sales manager experience. They could not estimate the elasticity of selling effort precisely. It could be increased by about 50% without changing the reported $R^2$ because of the flat response surface arising from the high correlations among the independent variables in their nonlinear model. Parsons & Vanden Abeele [1981] found the sales-call elasticity of an established Belgian ethical drug to be positive, but inelastic.

Little information exists on product and distribution elasticities because changes occur rarely, or slowly, for established products. Thus, their effects are usually represented in sales response functions as part of the constant intercept. Although these effects cannot be identified, this does not mean they are unimportant. For additional discussion of brand generalizations, see Erickson [1990]. Having seen what is known about brand sales, we now turn to industry sales.

5.2.2. Industry-level generalizations

Models of industry demand have also been constructed to assess the impact of trade-association or government efforts and to address public policy questions. For example, wool producers might want to determine the effectiveness of advertising the 'Wool Mark'. In the same vein, public-health officials might want to evaluate the relationship between cigarette advertising and children's cigarette consumption. These would be examples of 'primary' advertising.

Nerlove & Waugh [1961] discovered that the advertising and promotion expenditures of the two largest organized groups of growers, the Florida Citrus Commission and Sunkist Growers, had a marked impact on the sales of oranges in the United States. Ball & Agarwala [1969] determined that generic advertising for tea in the United Kingdom slowed the downward sales trend, but could not reverse the slide. McGuiness & Cowling [1975] decided that advertising had a positive, statistically significant impact on cigarette sales in the United Kingdom, and that this impact was only partially offset by the amount of publicity given to the health effects of smoking. Lambin [1976] found that in only four out of ten product markets did industry advertising increase industry sales. Simon & Sebastian [1987], using a model with systematic parameter variation, showed that advertising influenced the diffusion of new telephones in West Germany. Their advertising (goodwill) elasticity attained a maximum of 2.14%, then declined nonmonotonically to 0.89%, within five years. These and other studies lead to the generalization that primary advertising has a direct and positive influence on total industry (market) sales [cf. Leone & Schultz, 1980].

Industry sales are less price-elastic than individual brand sales. Industry price-elasticities are typically less than one in absolute magnitude (inelastic) whereas brand price-elasticities are typically larger than one in absolute magnitude (elastic). Neslin & Shoemaker [1983] found that the presweetened segment of the ready-to-eat cereal market in the United States was much more price-sensitive than the market as a whole. If industry elasticities are less than one (inelastic), then why
do firms not raise price? The answer is they do when they can—such as in the case of oligopolies with a leader, e.g., ready-to-eat cereal with Kellogg. With what we know now, where might we go next with our research agenda?

6. Research issues

While we know quite a bit about advertising and something about price, our empirical base on other marketing instruments is woefully small. A better grasp of marketing instruments other than advertising is needed. For instance, marketing managers would like to know when and to what degree excessive price-promotion changes customers' perceptions of the normal price of a product and of its quality. Also, managers would like to know how to assess the impact of a salesperson's effort, especially since empirical evidence has indicated that territory workload and potential are significant determinants in sales differences among territories, whereas sales effort may have little, if any, effect. Ryans & Weinberg [1979] conjecture that it might be useful to construct a two-stage model of sales-force performance. The first part would specify the factors that influence the amount of effort a salesperson puts forth and the second would represent the relationship of sales to sales-force effort.

Some empirical results, such as those involving the product life-cycle, seem to be in conflict with marketing theory. Other empirical results, such as those for the price—advertising interaction, have been contradictory. Consider, for example, the relationship between the demand elasticity and relative price. Moran [1978] states that the farther a brand's price is from the category average in either direction, the lower its demand elasticity is, whereas Simon [1979] says that the magnitude of price elasticity increases for increasing positive and negative deviations of a brand's price from the average price of brands competing with it. Parker [1992] found no consistent pattern of price-elasticity dynamics over the adoption life-cycle for 19 consumer durables. We will now focus our attention on major emerging trends in ETS.

6.1. Emerging trends in econometric modeling

While the use of econometrics in marketing is now mature, nonetheless there are exciting prospects for new advances. Major developments relate to definition of variables, under-researched mix elements, model identification, Bayesian estimation based on meta-analysis, combining different functional forms, level of aggregation, discretization of data, and efficiency frontiers.

6.1.1. Definition of variables

How should we measure marketing instruments properly? For example, in many, but not all, sales-response models, 'advertising' is measured in monetary units, e.g., dollars, but consumers do not react to dollars. In a similar vein, is an unpromoted temporary price-cut a 'promotion'? How temporary should a price-cut
be to be classified as a 'promotion'? Should price be price/weight, or price paid, or price/delivered benefit (e.g. price/cleaning power) for detergents? These and related questions need to be addressed to refine our assessments of the effects of marketing instruments.

6.1.2. Under-researched mix elements

While advertising and price have been widely studied, and price promotion is currently under intense scrutiny with the availability of scanner data, distribution and sales-force effort have been little studied. Farris, Oliver & De Kuyver [1989] propose an interesting approach to modeling the relation between distribution and market share. Total market share is broken down into share due to uncompromised demand and share due to compromised demand from buyers whose preferred brand is unavailable. These two share variables are then modeled as different functions of distribution. Distribution is made operational by using PCV, the weighted (by-product-category volume) fraction of stores stocking the brand. (They note that many retail audit services report ACV, which is weighed by all-commodities volume; and that these two measures (PCV and ACV) may not be equal.) More research on this approach, and distribution in general, is needed. Very few studies have been able to separate out the effects of sales-force effort from the overwhelming influences of potential and workload. Yet sales-force effort is often crucial to the successful marketing of products and services.

6.1.3. Model identification

We do not want to leave the false impression that you only need to grab some data set to obtain parameter estimates of the functional form of your model. Econometric techniques explain the impact of variation in the explanatory variables on a dependent variable. But what if an explanatory variable does not vary much? For example, a brand manager of the leading brand in the soap market knows that if he increases price, brand unit sales will go down dramatically. If he decreases it, he will lose money. Consequently, the price of the brand is likely to remain relatively constant over a several-year time horizon. What can an econometrician say about the price parameter in this circumstance? Very little. We need to know the conditions to be met for a model to be identifiable from the data.

6.1.4. Bayesian estimation based on meta-analysis

Farley & Lehmann [1986] believe that Bayesian applications of meta-analysis are a promising area for research when data are sparse or when preliminary estimates are needed such as in the case of new products. The results from meta-analysis can be considered prior estimates; and then Bayesian regression can be used to get modified estimates for the market-response function. In this spirit, Russell, Hagerty & Carman [1991] apply a Bayesian methodology to estimate firm-level marketing-mix elasticities when sparse data prevent classical regression procedures from recovering elasticity estimates. Elasticity estimates for over 2200 firms in the PIMS database are recovered by using the elasticities of 197 firms found earlier [Hagerty, Carman & Russell, 1988] as prior information.
6.1.5. Separate response functions

Most research on market-response functions assumes that each of the marketing-mix instruments follows the same response function. In particular, each instrument is usually assumed to follow a power function, which, in turn, can be combined multiplicatively with other instruments to directly form the multiplicative sales-response function. However, each instrument could follow a different functional form. In MARMIX, a computer-based model-building package designed to help managers in setting annual marketing budgets, De Kuyver & Pessemier [1986] permit separate response functions for the marketing-mix elements to be estimated. These separate functions are then combined in an additive model if independence among the different elements in the marketing mix is assumed, or in a multiplicative model if interaction among these elements is expected. Moore & Pessemier [1992] discuss more recent refinements of MARMIX. Work in this direction has just begun.

6.1.6. Level of aggregation

The 'market' in 'market response' refers to a group of heterogeneous individuals. One problem that has not been addressed is the lack of recognition of individual heterogeneity in a changing population over time. Moreover, while we have discussed store-level versus market-level aggregation, nothing much has been said about the individual level where response actually occurs.

We do want to call attention to econometric developments as they relate to the total number of purchases by a customer in a particular time period. Purchase incidence has been modeled as a Poisson-distributed variable, in which heterogeneity among customers is captured by allowing their purchase rates to vary following a gamma distribution. Parsons [1987] and Rosenquist & Strandvik [1987] have proposed representing heterogeneity directly by relating explanatory variables to purchase incidence using Poisson regression:

$$
Pr\{N_t = k|\lambda\} = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \ldots
$$

(6.1)

where $N_t$ is the number of purchases at time interval $t$ and $\lambda$ is a parameter.

Explanatory variables are introduced by making the parameter of the Poisson distribution, $\lambda$, a function of them:

$$
\lambda = f(X, \alpha)
$$

(6.2)

where $\alpha$'s are the parameters. This is but another example of (deterministic) systematic parameter variation. The functional form of this relationship is usually linear or log-linear, but could be nonlinear. Estimation is usually done by maximum likelihood estimation (MLE) based on the Poisson distribution, but Bayesian and pseudo-MLE methods have been put forward as well. The Poisson distribution has been criticized because of its property that the variance equals the mean:

$$
\lambda = E(N_t) = \text{Var}(N_t) > 0.
$$

(6.3)
Extensions have been proposed to relax this assumption. Modified count data models include *hurdle models* and *with-zeros models*.

Despite the research noted earlier, more work needs to be done on the impact of aggregation on the functional form of the response function and the parameter estimates. Areas of interest include individual- versus aggregate-level estimation, weekly versus monthly estimation, cross-section versus time-series versus cross-section-and-time-series estimation.

6.1.7. Discretization of data

The process being modeled is inherently continuous, yet we usually build models based upon discrete data. A number of the response phenomena might be artifacts of this discretization.

The main issue in the specification of a dynamic model is whether to formulate a continuous-time model using differential equations, a discrete-time model using difference equations, or a mixed model using differential–difference equations. Before discussing this issue, it is necessary to make a distinction between instantaneous variables and flow variables. An *instantaneous (stock) variable* is a variable that is measurable at a point in time. Prices, distribution coverage and interest rates are examples of instantaneous variables in marketing. A *flow variable* is a variable that is not measurable at a point in time. Unit sales and advertising expenditures are examples of flow variables in marketing. The rate of change in an instantaneous variable is also treated as a ‘flow’ variable.

The primary advantage of a continuous model is that its *estimated parameters are independent of the observation period*. This does not hold for a discrete model. A discrete model estimated on weekly data will be different from one estimated on annual data. The two primary advantages of a discrete model are that it can capture discontinuous phenomena and that most econometric techniques have been developed for estimating it. Some work on this topic has been done by Houston & Parsons [1986] and Rao [1986].

6.1.8. Efficiency frontiers

We have discussed various specifications of sales-response functions and have noted the presence of equation error. Random disturbances are assumed to occur around the estimated sales-response function. These errors occur on both sides of the estimated function. Thus, the traditional econometric approach may be considered to produce an ‘average’ sales-response function. This is sufficient for many purposes. However, in other cases such as that of marketing productivity, it is inadequate.

Consider sales-force productivity. Econometric methods are often used to either set quotas or to assess the performance of a salesperson in a territory in order to take into account the many factors, such as potential and workload, involved. A sales manager does not want to use ‘average’ performance as a standard, but rather wants people to strive to achieve the best that can be achieved in their territories. Therefore, a need arises to estimate the sales-response function that expresses the ‘maximum’ sales achievable from the effort of the sales force given the environments
in which its members operate. This function is called a frontier sales-response function.

Observations will either lie on a frontier or fall below it. The result is a model with a one-sided distribution of errors:

\[ Q = f(X, \beta) - v \quad \text{where } v \geq 0. \tag{6.4} \]

These errors represent inefficiency. Distributions used include the half-normal, exponential, truncated-normal, and the gamma. Estimation is done by corrected least squares or maximum likelihood.

One-sided distributions, however, have their weaknesses. Measurement error in the dependent variable would be a major problem in the analysis. Moreover, as might be expected, outliers can dominate the estimated sales-response function. This has led some to add a traditional two-sided error term to the model as well:

\[ Q = f(X, \beta) - v + \omega = f(X, \beta) + v. \tag{6.5} \]

We now have a frontier model with composed error. For a recent review of econometric estimation of frontiers, see Bauer [1990]. Marketing researchers working in this area include Parsons [1991] and Horsky & Nelson [1991].

6.2. Emerging trends in time-series modeling

Although a lot has been written on distributed-lag relationships among marketing variables, we know remarkably little about the long-term effects of marketing actions on sales and other performance variables. A key reason for this is that our market-response models typically assume that the data are generated as stable (usually normal) distributions around a fixed mean. Therefore, if long-term movements in the data are present (such as an upward trend followed by a downturn), our models are not well equipped to represent their underlying causes. Addressing long-term effects is one emerging trend. Another one is the untangling of the effects of purchase reinforcement and advertising carryover.

6.2.1. Long-term effects

Recent developments in long-term time-series model-building, especially unit-root testing and cointegration modeling, offer substantial promise for gaining new insights on the long-run effects of marketing efforts. In this section we briefly review the key aspects of long-term modeling. For a complete application we refer to Powers, Hanssens, Hser & Anglin [1991].

Temporary and permanent components. It is often observed that marketing and other socioeconomic time-series move smoothly over time. Suppose such is the case for a hypothetical brand's sales \( Q \) and that a simple first-order Markov model or autoregressive process can be used to represent this phenomenon:

\[ Q_t = \phi Q_{t-1} + c + \omega_t \tag{6.6} \]
or, using lag operator notation,

\[(1 - \varphi L)Q_t = c + \omega_t,\]

(6.7)

where \(\omega_t\) is white noise. The time series of sales has the simple properties of a constant mean and a constant variance. Furthermore, successive substitution of lagged sales in Model (6.7) ultimately produces the expression

\[Q_t = c + \omega_t + \varphi \omega_{t-1} + \varphi^2 \omega_{t-2} + \cdots\]

(6.8)

in which current sales levels are explained as the sum of an infinite number of random shocks in the past. With \(|\varphi| < 1\), the effect on any shock \(\omega\) on \(Q\) eventually dies out, i.e. the time series \(Q_t\) consists of strictly temporary components. If the series drifts away from its mean, it must eventually return.

If \(\varphi = 1\) the situation changes drastically. Now the infinite-shock representation of the model is

\[Q_t = c + \omega_t + \omega_{t-1} + \omega_{t-2} + \cdots\]

(6.9)

and it is clear that any shock in the past, however distant, has a permanent effect on sales. The reduced form of such a model is

\[Q_t = Q_{t-1} + c + \omega_t\]

(6.10)

which is commonly known as the random-walk model, and may be 'without drift' \((c = 0)\) or 'with drift' \((c \neq 0)\). The properties of the random walk are that it has no fixed mean and that its population variance is infinite, i.e. the variance grows without bounds as the time series expands. In contrast to a temporary-effects model, this model does not imply that the series must return to any previous value. It may wander in any direction for as little as one period or as many as one hundred or more periods. In fact, the expected time of return to any observed value is infinite. The best-known example of a random walk is the behavior of stock-market prices.

In practice we may observe time series with a mixture of temporary and permanent components. The statistical distinction between the two involves the presence or absence of unit roots in the characteristic equation of the time series. For example, in the model above, the characteristic equation is

\[1 - \varphi z = 0,\]

(6.11)

which has a unit root only if \(\varphi = 1\). Consequently, testing for the presence of permanent components in time series is equivalent to testing for the presence of unit roots in the time-series model underlying the data. Such tests are far from trivial, because the statistical properties of the data are different in the presence as opposed to the absence of unit roots. See Dickey, Bell & Miller [1986] for a comprehensive review.
Implications of unit-root testing. Unit roots in a time series of brand sales imply that the sales fluctuations are more than just temporary deviations from a constant mean. For example, sales may grow for long periods of time, and at different rates of growth, they may suddenly peak and even take on a downward course, as would be typical of a product life-cycle. These developments are typically not deterministic, for example, a growth trend may stop at any time, and it is precisely this type of sales evolution that is realistic in the marketplace and of obvious managerial relevance. On the other hand, stationary data (without unit roots) are perhaps less interesting since the long-run outlook of sales in a stationary system is known from the sample.

Unfortunately, it is difficult to build models on data with unit roots. Since the population mean is undefined and the variance is infinite, traditional statistical inference runs into problems because the basic assumptions of the general linear model are violated. Time-series analysts solve this problem by transforming the data prior to model-building: first-order differencing removes linear trend, second-order differencing removes quadratic trend, and seasonal differencing removes seasonality. Granger [1980] refers to the original series as \textquote{integrated of order 1}, or \textquote{I(1)}, and the differenced series as \textquote{I(0)}. In general, if a time series requires \textit{kth}-order differencing in order to become stationary, i.e if it has \textit{k} unit roots in its characteristic equation, then the series is said to be \textquote{I(k)}.

While differencing the data to obtain stationarity is common practice in time-series model-building, the resulting model in changes operates only the short run: temporary fluctuations in \textit{Y} can be traced to temporary fluctuations in \textit{X}. Such models do not make inferences about the long-run behavior of \textit{Y} in function of that of \textit{X}. Consequently, as soon as the forecasting horizon is extended, the predictive power of such models drops considerably.

The equilibrium relationship. Several literatures, most notably economics, statistics, engineering and management, use concepts such as \textquote{equilibrium}, \textquote{steady state} and \textquote{long-term relationship}. Although individual definitions may vary somewhat, we can say, in general, that in an equilibrium relation between \textit{X} and \textit{Y}, there will be no change in either \textit{Y} or \textit{X}; furthermore, there will be a relationship \textit{Y} = \textit{cX}. Formally:

\begin{align*}
  Y_t - Y_{t-1} &= 0, \quad (6.12a) \\
  X_t - X_{t-1} &= 0, \quad (6.12b) \\
  Y_t &= cX_t, \quad (6.12c)
\end{align*}

or, in words, there are no pressures for either sales or marketing to move away from present levels, and there is some constant ratio between the two. In practice, of course, there will be random perturbations from equilibrium, but they will be strictly zero-order. For example, the first-order differences of the data may be white noise:
\[ Y_t - Y_{t-1} = (1 - L)Y_t = \omega_{yt} \]  
\[ X_t - X_{t-1} = (1 - L)X_t = \omega_{xt} \]  

(6.13a)

(6.13b)

The steady-state relationship between \( X \) and \( Y \) prevents the individual series from wandering away too far from each other. As a simple example, consider the relationship between production levels and retail sales. The equilibrium between the two is maintained by an inventory position which is acceptable to both parties in the exchange (the marketer and the consumers). If, for some external reason, retail sales suddenly start to move upward, then inventories are depleted and production must eventually catch up. Thus, while individual series may well be nonstationary (sales in the example), equilibrium implies that the (scaled) difference between the two is not allowed to permanently deviate from zero.

We are therefore in a position to formally incorporate an equilibrium relationship in a market-response model or other system. Following a recent definition by Engle & Granger [1987], two series \( X \) and \( Y \) are said to be cointegrated if they are each \( I(k) \), but there exists a linear combination of the two series which is stationary:

\[ Z_t = X_t + cY_t \sim I(0), \]  

(6.14)

where \( c \) is called the integrating constant. The definition is conceptually appealing. In the absence of cointegration, any arbitrary linear combination of two non-stationary time-series will also be nonstationary. Also, in the seasonal time-series that are characteristic of many marketing situations, it will typically be the case that the combination of two seasonal series is itself seasonal. But if the series are cointegrated, then a powerful result emerges, namely that the (scaled) deviation between the two is stationary. If more than two series are involved, the concept may be extended to that of a cointegrating vector.

Testing for an equilibrium. The definition of cointegration is not only intuitively appealing, it also allows for empirical testing on time-series data. Having established, for example, that \( X \) and \( Y \) are each \( I(1) \):

\[ \Delta Y_t = \omega_{yt}, \text{ so } Y_t \text{ is } I(1), \]  

(6.15a)

\[ \Delta X_t = \omega_{xt}, \text{ so } X_t \text{ is } I(1). \]  

(6.15b)

If \( X \) and \( Y \) are cointegrated, then the time series of residuals,

\[ u_t = Y_t - bX_t, \]  

(6.16)

should be \( I(0) \). Engle & Granger [1987] propose to estimate the 'equilibrium regression'

\[ Y_t = a + bX_t + u_t \]  

(6.17)
by ordinary least squares and to subject the residuals $\hat{\epsilon}$ to a Dickey–Fuller test. Rejection of the null hypothesis of nonstationary residuals is evidence that the time series are in fact cointegrated. A few comments about the 'equilibrium regression' are in order. At first glance it is a naive equation, estimated by a basic method (OLS) which does not reflect the state of the art in applied econometric modeling. The residuals are likely to be ill-behaved, for example, they may be highly autocorrelated. But the essence of the equation is not the common hypothesis test on $b$ or an overall goodness-of-fit test, but rather the challenge of taking two or more nonstationary time-series and finding a linear combination of them that is stationary. Stock [1987] has shown that, under the hypothesis of cointegration, the OLS estimate has three desirable properties: it is consistent, it has a bias of order $T^{-1}$, and it has a variance of order $T^{-2}$. Thus for all but unduly small samples the OLS estimate provides a reliable estimate of the cointegrating vector.

In conclusion, we may formulate two conditions for the existence of a long-run or equilibrium relationship between sales levels and marketing support. First, the sales series must contain at least one permanent component. Secondly, these permanent components must be related to each other. If these conditions are not met, we may still observe strong relationships between the series, including distributed-lag relationships, but these would only affect the temporary behavior of the series. For example, sales promotion may increase market share in the same period and decrease it in the next period (due to consumer stockpiling), but eventually sales will return to their average levels. However, if an equilibrium relation is present, then the use of sales promotions may eventually cause market shares to grow to higher levels.

To date, cointegration models have been used in marketing to assess the long-term market structure of the private aircraft industry [Dekimpe & Hanssens, 1991] and to measure the long-run impacts of public demarketing efforts to curb narcotics abuse [Powers, Hanssens, Hser & Anglin, 1991]. Several other areas of research in marketing stand to benefit from this approach, for example, product life-cycle theory and brand-equity research.

6.2.2. Untangling purchase reinforcement and advertising carryover

We have already noted that at least some of the impact of advertising in one time period may be carried over into future periods. The argument is that past advertising is remembered by those who see it and, as a result, a brand builds 'goodwill'. This goodwill influences brand choice. This approach may be too simplistic in that it ignores purchase experience. Advertising may create the initial purchase, but customers will buy a brand again only if they find it acceptable in use. Thus, an alternative to an advertising carryover approach is a current advertising effects with purchase feedback approach. In reality we would expect both to be operating. Thus, Givon & Horsky [1990] propose a model to estimate simultaneously the relative magnitude of the two approaches. Their empirical work indicates that purchase reinforcement dominates over advertising carryover in affecting the evolution of market share. Dekimpe & Hanssens [1993] propose a new measure called 'persistence' to assess these effects in the long run.
7. Discussion

Marketing has seen a rapid expansion in the widespread use of quantitative methods. Correlation and regression analysis were among the first techniques used as marketing research emerged as a discipline after World War II [cf. Ferber, 1949]. In the early 1970s regression analysis became econometrics. Simultaneous-equation systems could be estimated almost as easily as single regression equations [e.g. Bass & Parsons, 1969]. While econometrics as a whole continues to flourish as new and more sophisticated estimation techniques and associated computer software have become available, simultaneous-equation systems have not become widely prevalent. One explanation is that an applied researcher may find little difference in practice between parameter estimates obtained from a technically correct simultaneous-equation technique and those obtained from a corresponding 'sinner' single-equation technique. Notwithstanding this possibility, allied social-sciences disciplines have experienced a recent resurgence in the use of simultaneous-equation models not yet seen in marketing. For example, simultaneous-equation models currently comprise more than 40% of the econometric articles in agricultural economics [Debertin & Pagoulatos, 1992, p. 8]. Perhaps what is being indicated is a need for more theory in marketing.

Knowledge about the nature of market mechanisms has increased appreciably. For example, not long ago managers had little but their own subjective impressions of the effectiveness of their advertising. Managers would say things like 'I know I am wasting half my advertising spending... I just don't know which half'. Today managers of heavily advertised brands can measure the short-run effects of advertising spending with reasonable accuracy. This research stream has progressed to the point where we now have a good handle on the average advertising elasticity of a frequently purchased branded good. Even managers of industrial products have at least a benchmark against which to evaluate their advertising effort [Lilien, 1979; Lilien & Ruzic, 1979]. Our task now is to understand the more subtle variations in marketing's impact.

Although the shape of the sales-response function is almost surely concave over realistic operating ranges, we should be alert to the possibility of other shapes. Johansson [1979], for example, has suggested one approach for identifying whether or not a relationship under analysis is S-shaped. In general, because different combinations of phenomena such as threshold and saturation effects might be present in a particular market, we should not think in terms of a single aggregate response function. A specific response function should be constructed for each product-market situation. A table of response functions and their characteristics is given by Doyle & Saunders [1990]. We present in Table 9.1 selected applications of the use of different response functions for each of the marketing-mix elements.

The research developments reviewed herein share an important common theme: market-response models are becoming closer to the behavioral and managerial realities of marketing. For example, the simplistic linear and logarithmic approximations buyer response to marketing are being replaced by more behaviorally realistic functions that take into account the managerial objectives of marketers.
Table 9.1.  
Selected marketing-mix applications of functional forms

<table>
<thead>
<tr>
<th>Functional form</th>
<th>Advertising</th>
<th>Price</th>
<th>Promotion</th>
<th>Sales force</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplicative</td>
<td>Bass &amp;</td>
<td>Popkowskij,</td>
<td>Parsons &amp;</td>
<td>Vanden [1981]</td>
</tr>
<tr>
<td></td>
<td>Parsons [1969]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lambin [1976]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rao [1990]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wildt [1977]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simon [1982]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>Cowling &amp;</td>
<td>Blattberg &amp;</td>
<td>Wisniewski</td>
<td>Bolton [1989] [1989]</td>
</tr>
<tr>
<td></td>
<td>Cubbin [1971]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-reciprocal</td>
<td>Metwally [1980]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bemmar [1984]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified</td>
<td>Shakun [1965]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exponential</td>
<td>Holthauer &amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assmus [1982]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic</td>
<td>Johansson [1973]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocal</td>
<td>Ward [1975]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translog</td>
<td>Jagpal, Sudit &amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vinod [1982]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplicative</td>
<td>Jagpal, Sudit &amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonhomogeneous</td>
<td>Vinod [1979]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jagpal [1981]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As data-processing and parameter-estimation technology evolves, implementing these more sophisticated models becomes easier.

Another emerging trend may have even more far-reaching implications. As the environment of marketing becomes more information-intensive, a new marketing strategy evolves around information technology [Glazer, 1991]. For example, many service organizations such as banks, credit-card companies and airlines use their extensive customer databases as strategic assets to develop cross-selling and other customer-loyalty-focused marketing strategies. The sheer size of these databases necessitates the use of market-response models in order to understand the driving forces of purchasing behavior and to predict the likely outcomes of alternative marketing strategies. For example, financial institutions need to know which factors contribute to customer loyalty and how promotional strategies can enhance this loyalty. Therefore, market-response models are becoming an inherent part of marketing strategy. We expect continuing managerial payoff of the ETS approach in marketing.
References


Russell, G.J. (1988). Recovering measures of advertising carryover from aggregate data: The role of the firm’s decision behavior. Marketing Sci. 7 (Summer), 252–270.


econometric models. Working paper, Johnson Graduate School of Management, Cornell University, Ithaca, NY.