PromoCast™: A New Forecasting Method for Promotion Planning

Lee G. Cooper • Penny Baron • Wayne Levy • Michael Swisher • Paris Gogos
The Anderson School, 110 Westwood Plaza, University of California at Los Angeles, Los Angeles, California 90095-1481, lee.cooper@anderson.ucla.edu
efficient market services, inc., 106 Wilmot Road, Suite 400, Deerfield, Illinois 60015
penny_baron@emsinfo.com • wayne_levy@ems.com • mike_swisher@ems.com • paris_gogos@ems.com

Abstract

This article describes the implementation of a promotion-event forecasting system, PromoCast™, and its performance in several pilot applications and validity studies. Pilot studies involved retail grocery chains with 95 to 185 stores per trading area. The goal was to provide short-term, tactical forecasts useful for planning promotions from a retailer's perspective. Thus, the forecast system must be able to handle any of the over 150,000 UPCs in each store's item master file, and must be scalable to produce approximately 800,000,000 forecasts per year across all the retailers served by *efficient marketing services, inc.* (*ems, inc.*). This is a much different task than one that confronts a manufacturer, even one with a broad product line. Manufacturers can benefit from custom modeling in a product line or category. Retailers need a production system that generates forecasts that help promotion planning. Marketing scientists have typically approached promotion analysis from the manufacturer's perspective. One objective of this article is to encourage marketing scientists to rethink promotion analysis from a different perspective.

From the retailer's point of view the "planning unit" is the promotion event. Neither weekly store-tracking data nor shopping-trip data from consumer panels are easily aggregated to reflect total sales during a promotion event. We describe the promotion-event databases and the statistical model developed using these databases. The data are the strategic asset. Our goal is to help retailers use their data to increase the profitability of promotions. We have data on the performance of each UPC in each store under a variety of promotion conditions, on each store's adeptness at executing various styles of promotions, as well as on chain-wide historical performance for each UPC. We use many historical averages from these databases to build a 67-variable, regression-style model. The forecast incorporates a simple bias correction needed when using a log-transformed dependent variable (the natural log of total unit sales). We argue that the historical averages matching the planned ad and display conditions provide a benchmark superior to the widely used "base-times-lift" method. When aggregated into case units (the natural unit for product ordering), 69% of the forecasts in our first validation study were within ± one case compared to 39% within ± one case using the appropriate historical averages. We report the results of two over-time validity studies that reflect the value of our model for retailers. The limitations and implications of this planning tool for managerial decision making concerning stocking levels are discussed.

Whenever historical data are the strategic asset we face inherent limitations. Our model does not forecast new products. The forecast error increases when an existing product is promoted in a new way. Over 99.5% of the time, we have full data from which to create a forecast. However, with a database for a typical chain market containing over 20 million promotion events in the 30-month time frame we use, 100,000 events have less than ideal data. The breadth of the database (typically 150,000 UPS) makes it impractical to incorporate data on competitive offerings. We find that regression-style modeling is not adept at incorporating information on the 1,200 subcommodities managed in our pilot stores or the 1,000 manufacturers who supply those stores.

Despite these limitations we show the value of using promotion-event data, how tactical forecasts based on these data can directly impact the bottom line of grocery retailers, and how store-by-store forecasts can help retailers with problems of running out of stock or overstocking.
(*Retailing; Promotion Planning; Forecasting; Promotion Event Data*)

0732-2399/99/1803/0301/$05.00
1526-548X electronic ISSN

Introduction

Promotion planning is a daily task for both manufacturers and retailers in the consumer packaged goods (CPG) industry. The 3,200 grocery retailers tracked by *ems, inc.* in September 1997 sold an annualized volume of over $48 billion worth of goods, with $13.3 billion coming from sales on various promotions.[1] The planning task for retailers is very different from that of manufacturers. A broad line for a manufacturer may have hundreds of UPCs that could be promoted. This is small compared to planning for the 30,000 items that are in stock at any given time for a retailer, or the 150,000 UPCs maintained in the retailer's item master file, from which those 30,000 are selected. Being able to plan promotions for any of these 150,000 UPCs is a far more daunting task than any manufacturer faces.

The rapid diffusion of scanner technology over the past two decades has helped foster the belief that this vast information resource can be harnessed to make accurate promotion planning a routine endeavor. So far, the dream has far exceeded the reality. Anderson Consulting[2] estimates that 15% of promoted volume is lost due to out-of-stock (OOS) conditions. The issue is high on customers' agendas. After "good prices" and "fast/efficient checkout," customers most desire the store to be "in stock on the merchandise they want." When out-of-stock occurs, 20% of customers report leaving the store without purchasing any merchandise.[3]

The problem is exacerbated by current industry practice concerning ordering inventory for promotion events. In recent interviews with executives from seven leading grocery chains, the most sophisticated practice, for an upcoming promotion, was to order the quantity that was ordered chain-wide (in the trading area) for the "last like" promotion. "Last like" promotion was defined as the last time a promotion for the item was offered at the same price point. This practice has many downsides. First, no other promotion conditions (type of feature or display) were considered in establishing "last like." Second, "price" has too many levels to be a practical matching criterion. Many ad hoc rules have to be created to determine what is considered "the same price." More importantly, note also that the new order quantity relates to the last order quantity, not last sales (plus safety stock). This characterizes a system that does not learn; as a result, over-ordering (or under-ordering) on the prior promotion is repeated. The allocation of a chain-wide order to the stores in the trading area is also a problem. The most common practice involves classifying stores into three or four store-size groups and allocating stock proportional to these sizes. Some chains grouped stores according to the size of the category involved in the promotion, rather than by overall store size. Again, the most sophisticated practice was to look at shipments (query a database) by "last like" promotion, but then allocate proportional to "last like" sales in each store.[4]

Even if management could accurately forecast aggregate sales for an upcoming promotion, the absence of a store-by-store forecast would be costly. To illustrate this, we will use the results from a recent promotion for Tide™. The store-by-store sales and the forecast values based on our model, when the promotion was run in 80 of the 95 stores in pilot market 1, are shown in Figure 1. The ability of the forecast to reflect store-by-store differences is clear, but this is the topic of later discussion. The pertinent issue here is the decision concerning allocation to stores. Allocating equal amounts of product to each store would be a poor policy, even if highly accurate chain-wide forecasts were available. Even assuming a perfectly accurate chain-wide forecast, over 2,700 units (34% error) would be misallocated across the 80 stores if based on chain-wide average sales.

Could we allocate products to stores on the basis of store size? If stores are grouped into four quartiles

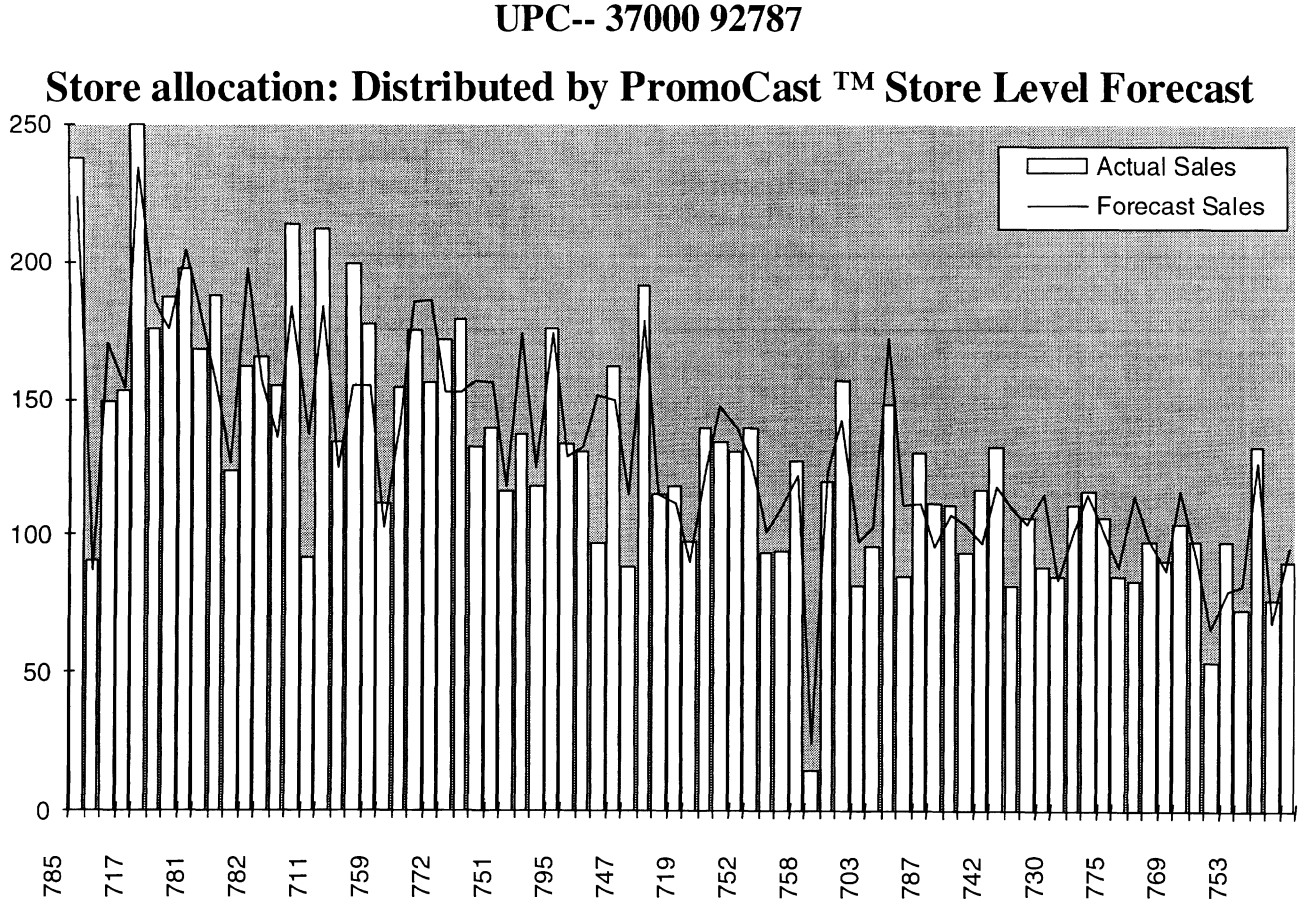
[1]This represents 10.6% of the supermarkets (with sales of $2 million or more) in the U.S. and 14.4% of the dollar volume sold in all supermarkets, according to the Food Marketing Institute 1998 figures.

[2]"Where to Look for Incremental Sales Gains: The Retail Problem of Out-of-Stock Merchandise," Anderson Consulting Study conducted for the Coca-Cola™ Retailing Research Council, 1996.

[3]Op. cit., Leo J. Shapiro and Associates conducted interviews showing that 20% of consumers continued their search at another store (and did not make any purchases), 20% delayed purchase, and those who purchased substitutes spent 6% less for the alternatives they bought.

[4]The most widespread industry practice, involving *base-times-lift* calculations is discussed later.

Figure 1 Tide with Bleach Powdered CP—March 1997



based upon total sales and products allocated proportionally—even assuming a perfectly accurate chain-wide forecast—over 2,160 units (27% error) would be misallocated across the 80 stores based on store size. If category size is used as a basis for a store allocation rule—again, assuming a perfectly accurate chain-wide forecast—over 1,940 units (24% error) would be misallocated across the 80 stores based on category size. Both of these policies would be greatly inferior to one based on allocation by the PromoCast™ forecast, for which 1,150 units (14% error) would be misallocated. Clearly, our model captures store-specific information for an item that is valuable in developing not just chain-level forecasts, but store-by-store allocation policies.

If we generalize the proportional error reduction in this example to the corresponding *profit* savings from a reduction in out-of-stock and overstock for a typical store, we can see what is at stake. If each store in the chain were allocated the chain average, our store-by-store forecast would increase profit by $369,000 per year in a typical 100-store chain. Compared to an allocation to each store based on store size, our store-by-store forecast would increase profit by $339,000 per year in a typical 100-store chain. Compared to an allocation based on category size, the store-by-store forecast would increase profit by $302,000 per year in a typical 100-store chain. These are conservative estimates that assume the grocery chain can accurately forecast chain-level results without a method such as the one to be discussed in this article. Systems that rely on principles such as ordering "last like" shipment amounts are far less accurate (as will be detailed in the Discussion section).[5]

Fundamental planning simply involves being able to do a short-term, tactical forecast of how much *each store* in a grocery chain's trading area will sell under a given combination of marketing efforts for an item. The development of appropriate market-response models depends on the databases available. The primary databases supporting the planning effort have been consumer panel data and weekly (or daily) store-tracking data (i.e., a shopping trip). The temporal unit of analysis for a consumer panel is the household visit to a store. While this unit of analysis has driven the extensive development of discrete choice models by academics in marketing and geography, it is not the natural unit of analysis for promotion planning. From the grocery retailer's point of view, the unit of analysis for promotion planning should be the promotion event, not the shopping trip. Thus, the management concern is to figure out what to order for an upcoming promotion event. The argument is that promotion-event data provide a complete *census* of store data that directly apply to the managerial decision, while shopping-trip data provide *sample* data that have to be dramatically transformed to answer the simple question of how much product to order for an upcoming event. Shopping-trip data do not easily answer this question.

The temporal unit of analysis for store-tracking data is the week (or, in some cases, the day). Approximately 46.4% of all promotions last longer than one week. As such, a retail manager can query a store-tracking database, asking for the sales of a UPC in a last promoted week. However, the retail manager needs to know whether that "last promoted week" was the first, second, or third week of an ongoing event. In essence, one would have to construct a promotion-event database. Thus, despite the best efforts of many academics (including the first two authors), market-response models based on store-tracking data are temporally ill-suited for the task of promotion-event forecasting. With a great deal of effort, one can make store-tracking data look more like promotion-event data. Since promotion-event data are available there is no need to bother transforming store-tracking data.

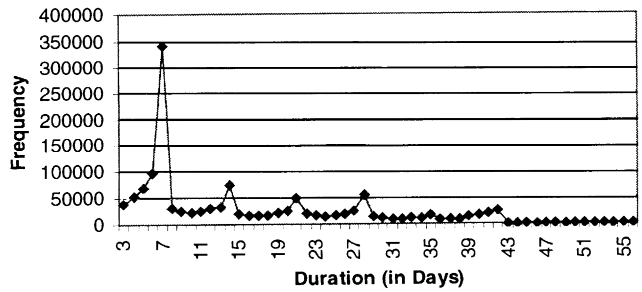
The proper temporal unit is the promotion event. Figure 2 shows the duration of over 1.3 million promotion events from a stratified random sample of promotion events used in calibrating our first pilot market.[6] We note that the biggest spike is at seven days, but additional spikes occur at 14 days (16% fall between 8 and 14 days), 21 days (9.6% fall between 15 and 21 days), and 28 days (20% of promotions are longer than 21 days).[7] Based on this chart and our knowledge of industry practice we can surmise that the primary planning periods for promotion events are 1, 2, 3, or 4 weeks in duration. Furthermore, the durations in-between most probably reflect some change

[5]To evaluate the costs associated with different allocation policies, we evaluated overstocking costs at the cost of working capital (12% per year). This is conservative since it assumes no cost for labor or storage space. The out-of-stock costs were estimated at 20% of the average market basket expenditure of $19.22 (as reported by the Food Marketing Institute) times a 15% gross margin. The 20% figure comes from the industry study cited above that showed consumers encountering an OOS condition left the store without making any purchases. To project the total store-profit loss, we took the estimated proportion of reported total store sales that were promoted sales as the *ems*-wide average of 27.7%. We then scaled the dollar volume of this sale up to the storewide total. These volumes were then scaled up to a typical 100-store chain in a trading area. Given the cost differences between overstocking and understocking, one might erroneously think that merely increasing the safety stock could solve the problem. This issue is addressed in the Discussion section.

[6]The promotion-event data are supplied by *efficient market services, inc.*

[7]Excluded from this figure are the one- and two-day "manager specials" that we believe to be unplanned events responding to overstocking or competitive deals. Also excluded are the 36 events with durations longer than 56 days.

Figure 2 **Duration of Promotion Events (N = 1,364,460)**



in the planned activity that triggers the end of a promotion event to the in-store computer.[8]

The unit of brand aggregation is also an issue. The retail business necessitates placing orders by UPC, rather than by a higher level of aggregation. In an anticipated promotion, a useful planning method must deliver forecasts unique to each UPC. The item master files of the stores in our test markets are typical of major grocery retailers, containing well over 150,000 unique UPCs in each of the pilot markets. In a two and one half-year time frame, each of our pilot markets had well over 20,000,000 promotion events, when counted as unique UPC by Store by Start Date events. The basic task is to forecast the results of a planned promotion for any of the over 150,000 UPCs in any store (within a chain), for any particular Start Date in the year, using the information available from over 20,000,000 historical promotion events in a retail chain in a trading area.

Method

The goal of this development effort is to use the information as a strategic asset in the promotion planning process. The obvious strategy is to ask, "What information is relevant to any promotion event?", and then to save and use that information. Three perspectives help us understand what information is relevant. The first view is of the promotional mix. Regardless of what item or what store is involved, we expect different results for different combinations of price-cuts, ads, and displays.[9] We take a database perspective and ask, "For the planned style of promotion, based on historic performance, what are our expectations?" Continuing from this perspective, the second view involves the item (UPC) itself, as viewed by its promotion history. How well has this item done when previously promoted in this particular way? How well has the item done (on average) on any style promotion in the past in the focal store and across the chain? How good are the historic baseline[10] sales for this item at predicting promotion results? The third view recognizes that each store can be differentially effective at implementing a particular style of promotion. What is each store's historic relation between a particular style of promotion and the resulting sales volume? These three perspectives lead to the specification of 67 variables in the regression model used in the forecast. The complete list of these variables and descriptions of the measurement procedures are given in the Appendix.

Throughout the model specification, we will see that historical averages are the strategic assets that enable tactical forecasting. We need to consider the implications of this strategy for how we evaluate forecast accuracy. First, in model development, we expect that using historical averages to "predict" promotion-event results will lead to relatively high correlations, simply because the events we are "predicting" are components of the averages we are using. We do not expect any particular event to dominate since, on average, there are 19 promotion events in each historical average for matching ad and display conditions for an individual item over the 30-month time frame (on average, an item is promoted every seven weeks). Second, in actual forecasting of future events, we expect (and hope) that the historical averages correlate relatively highly with these future-event results. We "hope," since relatively high correlations will reflect that the historical data are the strategic asset we need. However, if we do no better than the historical averages in forecasting future events, we are wasting a great deal of effort. One natural benchmark is how much better our forecasts are than the historical averages for matching ad and display conditions. We will first go through the model specifications and then use this benchmark to evaluate our tactical forecasting model.

[8]A promotion event is defined as a constant set of promotion conditions. Whenever a change occurs in price, display, or feature conditions, the current promotion event is closed and a new one is begun.

[9]Structural models often stop here and ask, "What is the duration of an advertising effect?" (cf. Lodish et al. 1995 or Leone 1995), or "Is there a permanent component to the boost from a particular promotion?" (cf. DeKimpe and Hanssens 1995). Our focus is on the tactical decision of how much to order for a particular promotion, rather than the strategic issues.

[10]The baseline sales measure is designed to reflect expected sales in a nonpromotion period. The *ems* algorithm is based on a weighting of sales yesterday, a week ago, and eight days ago, if all those days are nonpromotion days. This value is computed daily by *ems, inc.* for each UPC. Since it does not change during a promotion event, a single value is stored in the promotion-event database.

The Influence of Promotion Style

We expect the promoted sales to vary, by unit price, the percentage discount (PO),[11] and whether the promotion is an X-for-the-price-of-Y sale. We also have four main effects for ad types: "A-large," "B-medium," "C-small," and "P-coupon" and nine main effects for displays: "Store Front/Promotion Display," "Front-Aisle End Cap Display," "Rear-Aisle End Cap Display," "Free Standing In-Aisle Display," "Store Rear Display," "Mid-Aisle End Cap Display," "Side-Aisle End Cap Display," "Other Display," and "Secondary Location Display."

Rather than forming the two-way interactions of four ad types and nine display types, or the three-way interactions of ad and display and the PO variable, we grouped ads into major and minor ads and grouped displays into major and minor displays before forming interactions. The analysis of over 837,000 deals from many retail chains supported grouping ad types and display types as follows: Classify "A" or "B" ads as Major Ads and "C" (line) or "P" (coupon) ads as Minor Ads. Classify display types "Store Front/Promotion Display," "Front-Aisle End Cap Display," "Rear-Aisle End Cap Display," or "Free Standing In-Aisle Display" as Major Displays, and display types "Store Rear Display," "Mid-Aisle End Cap Display," or "Side-Aisle End Cap Display" as Minor Displays. Both the Duncan Multiple Range Test and the Student, Neuman, Keuhls Test supported these partitions.[12] Accordingly, the following promotion-style interactions are included in the model: Major Ad × PO, Minor Ad × PO, Major Display × PO, Minor Display × PO, Major Ad × Major Display, Major Ad × Minor Display, Minor Ad × Major Display, Minor Ad × Minor Display, Major Ad × Major Display × PO, Major Ad × Minor Display × PO, Minor Ad × Major Display × PO, and Minor Ad × Minor Display × PO.[13]

The Influence of an Item's Promotion History

An item's historic sales should be a strong indicator of what sales to expect. Four aspects of an item's history are relevant: average same-store sales of the item on matching ad and display conditions, average same-store sales of the item on all promotions, average chain-wide sales of the item on matching ad and display conditions, and average chain-wide sales of the item on all promotions. We also included the interaction of the PO variable with the average same-store sales of the item on matching ad and display conditions.

We cannot use current baseline sales to predict promoted sales since such information is not available when promotions are planned (usually at least 13 weeks ahead of the event). But we do have average baselines from prior promotion periods. We use the average same-store baseline sales for matching ad and display conditions and the average same-store baseline sales on all promotions. We also use the average chain-wide baseline sales for matching ad and display conditions and the average chain-wide baseline sales on all promotions.

The Influence of a Store's Promotion History

We expect stores to vary in their response to different promotion styles, reflecting both managerial effectiveness and the demographic character of the store's trading area. A store's average sales across all items on promotions that match the planned promotion should be a good indicator of expected sales. This average is included in the model. We also use the store's average baseline sales across all items on promotions that match the planned promotion.

A review of the residuals from preliminary models also revealed that, when an item's average promotion-event sales in a store are substantially better (or worse) historically than its chain-wide average, residuals were large. In these cases, as historical average sales increased, so did the residuals. Consequently, we added indicators of these conditions and the interactions of

[11]We use the Blattberg and Neslin (1990) version of a Percentage-Off variable as defined in the Appendix. This is called PO in the model.

[12]These tests were run separately first, assuming feature only accounted for sales and then assuming display only accounted for sales. This is a relatively crude approach, but is better justified when looking at promotion policies across retailers than it would be within a single retailer.

[13]Note that there is a promotion type with no ad and no display. Such events must have a temporary price reduction (TPR) of at least 3% or the event would not have been coded as a promotion. Only price discounts of 3% or more trigger a TPR-only promotion event. This 3% level was determined by an *ems* survey of what clients would accept, as well as price sensitivity analyses.

them with the PO variable. We asked, "Is the store-item movement we are forecasting historically better than the chain-wide average for this particular item promoted in this particular way?" If the answer was yes, we set an indicator variable SGTC (Store Greater Than Chain) to 1.0 (0 otherwise), and formed the interactions of SGTC with both the same-store average sales on matching ad and display conditions and the chain-wide average sales on matching ad and display conditions. We created a corresponding indicator variable and set of interactions when the store was much worse than the chain on matching ad and display conditions for an item.

The review of residuals also revealed a within-store pattern. When the proposed promotion was historically better (or worse) than the average across all promotion types for a particular item, the residuals were large. Accordingly, we created indicator variable ISB that switches on when the proposed promotion was historically much better than the item's same-store average across all promotion types. We also formed the interaction of ISB with the item's average same-store sales on matching ad and display conditions and the interaction of ISB with the item's average same-store sales across all promotion types. We created a corresponding indicator variable and set of interactions when the promotion style was much worse than average.

Seasonality

While we expect numerous categories of items to possess distinct seasonal patterns, the active management of retail outlets seeks to counterbalance these seasonalities across categories as a means to smooth out cash flow. Thus, applying an overall seasonality adjustment to an item's sales history, without regard to the patterns arising within a category, would tend to create complex time-series problems where none may previously have existed (Franses 1995, Franses et al. 1995, Franses and Paap 1995). Our seasonality adjustment operates at the subcommodity level. A five-week centered moving average (exclusive of holiday periods) of total dollar sales in the subcommodity is used in the model to reflect the seasonal trend. This five-week centered moving average is computed across years as follows: Each week in a 130-week frame is labeled as to the week of the year (e.g., the 17th week of year 1 or the 20th week of year 2). A five-week frame for the 19th week, for example, would be weeks 17, 18, 19, 20, and 21 from years one, two, and possibly year three. As such, up to 15 weekly cumulative sales in a subcommodity[14] could be included in the average. A component was dropped from this series if it was a holiday period for a particular year. In the end, we have for each subcommodity a seasonal ACV number that has 52 values (one for each week of the year) and reflects how much dollar volume that subcommodity sells in a five-week frame over the prior two or three years.

The approach to special seasonal influences used in this forecasting model involves highlighting (with indicator variables) specific weeks when special events tend to boost sales: Super Bowl, Easter, Memorial Day, Labor Day, 4th of July, and Key-Pay weeks. The special character of Thanksgiving and Christmas led us to exclude these periods from the first release of our model.

Model Structure

The basic form of the statistical relation should be log-log in many price and unit-related variables and log-linear in the variables reflecting the combinations of ads and displays (Cooper and Nakanishi 1988, Blattberg and Neslin 1990).[15] The dependent measure is the (natural) log of the number of units sold in a promotion event. Since this model was designed to be transported (and recalibrated) across many grocery chains in many trading areas, the functional form of the variables was left as general as possible. The PO variable (and its interactions) were always put in log form. The display and feature variables were binary (and thus always in linear form). The historical averages were typically represented in both log form and

[14]The "subcommodity" is a relatively specific categorization of store geography, such as gelatin, dinner helpers and prepared meals, pudding, nectar fruit juices, fruit roll snacks, luncheon meats, popcorn, and yogurt. The chain in pilot market 1 maintained 1,200 separate subcommodities to classify all its SKUs.

[15]We note that Christen et al. (1997) investigated the potential bias when arithmetic averages are used in nonlinear models. We feel that the longitudinal validity studies reported later demonstrate that bias is not a problem in this application. The lack of observed bias is most likely helped by the stationary condition of the markets we investigated. In periods of price inflation or other sources of dramatic change, the potential for bias exists.

linear form, except where colinearity analysis (Belsley et al. 1980) indicated that having both forms was redundant.[16]

While ordinary least squares (OLS) procedures provide unbiased estimates of the parameters in models with logs as dependent measures, the means are estimated with bias. The average of the logs reflects the geometric mean rather than the arithmetic mean of the underlying sales units. This problem has been discussed in the statistical literature for over 50 years (Finney 1941, Neyman and Scott 1960), but still is little discussed in the marketing science literature, with the notable exception of Wittink et al. (1987). John Totten and Ross Link discussed fixes for this bias on the e-mail list amodlmkt@ewebcom.com. The simple bias correction appropriate for large sample applications such as ours comes from Miller (1984). In all forecasts reported here, we add one-half the mean-squared error to the estimate before being exponentiated back into sales units. The more elaborate procedure involves modification to the hat matrix, but only affects the sixth decimal place in the forecast of units. Our procedure is analogous to that used by Wittink et al. (1987) and by Totten and Link.

Slow Moving Items

The reality of promotions is that some items move quite well when promoted, while others do not. Slow-moving items simply do not respond as strongly to the promotional effort as do faster-moving items. To avoid the damping of promotion sensitivity that might result from using a single model to reflect such a heterogeneous response pattern, we decided to calibrate different parameters for slow movers versus standard movers. Across all promotions in our first test market, the median sales volume was approximately 2.7 units per day. The distribution is very positively skewed (the average units per day is 8.8).

The situation was even more extreme with our earliest developmental database. To take our first look at the prospects for forecasting, we drew a sample containing all the promotions for 3,000 items in a 30-month time frame (rather than a sample of promotion events, as described later), for which we computed the needed averages for only these items. We learned that such a sample grossly overemphasizes slow-moving items. The average promotion units per day in this sample was 2.9, and the median was 1.3. We took this median as the break point between slow-mover models and standard-mover models and calibrated separate parameters for promotion response for each pool. This led to an approximately 16% reduction in average absolute percentage error in units (APEU) during the developmental stage of this modeling effort. The bifurcation was about 8% better than a trifurcation explored at the same early stage of development.[17]

The Effect of Duration

We expect to sell more on a promotion of longer duration. Part of this expectation is due to an implicit constancy in our mental simulation. We think that if Brand X is on sale for two weeks, we expect to sell more than if Brand X is on sale for just one week. In reality, the brands promoted for one week are different than the brands promoted for two weeks. Thus, the mental simulation we should run is slightly different. Should Brand X sell more in two weeks than Brand Y sells in one week? Obviously, when we frame the question so as to perform the proper mental simulation, we do not know the answer. The policy may well be to keep an item on promotion (particularly on displays) until inventory runs out. This policy would mean that a lot of longer duration events would have slower movers. The point is that managers typically run an inappropriate mental simulation when asked questions about duration.

The same uncertainty can occur when speculating about the effects of ads or displays. If you ask managers if they expect to sell more on a Major Ad than a Minor Ad, the answer will undoubtedly be, "Yes." Since different items are put on Major Ads than on Minor Ads, we cannot be sure a priori of what to expect. Managers should check the data to see if they are performing the right mental simulations. Figures 3, 4, and 5 show the average units sold per promotion event

[16]The condition indices for all final models are less than 100, indicating no serious problem due to collinearity. Collinearity analysis was used for similar purposes by Bridges et al. 1995.

[17]We learned not only to change our sampling scheme, but also that APEU is an error measure that is too sensitive to errors in slow-moving items as is discussed later.

Figure 3 Test Market 1 Average Units with Major Ad

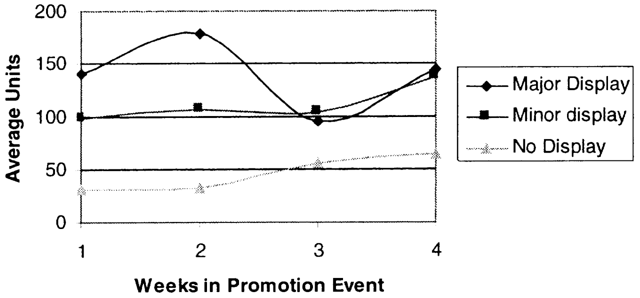


Figure 4 Test Market 1 Average Units with Minor Ad

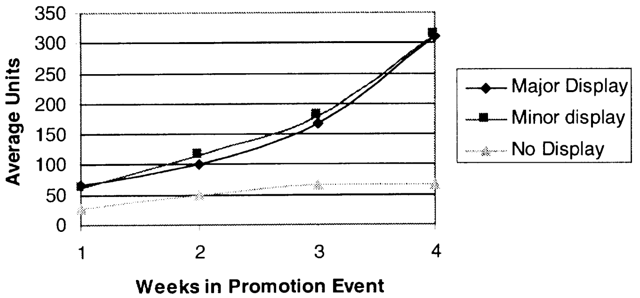
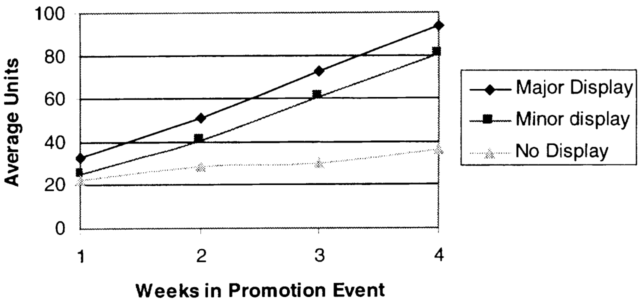


Figure 5 Test Market 1 Average Units with No Ad



for events of different durations in the major promotion conditions. We see in Figure 3 that longer promotions (three or four weeks) for Major Ads and Major Displays sell fewer total units (on average) than two-week promotions. This seems strange because we normally think about the expected effect of marketing instruments on sales. What we are actually seeing is the impact of a retailer's policy to promote certain (fast-moving) items using Major Ads and Major Displays for one or two weeks, but promote slower-moving items for longer periods. We expect this pattern to change from retailer to retailer and from one geographic market to another. Retailers have policies on how they promote and which items they place on what style of promotion. Trading areas differ in taste and receptivity to different styles of promotion. Even national retailers adapt their promotion styles to the regional character of each market. We should look at the data before we form our expectations.

Because of the complex relation of promotion duration to the other promotion effects and because of the interaction of policy variables (as compared to consumer-response variables), we decided to create separate promotion-response models for each of the four primary planning periods (1-, 2-, 3-, or 4-week durations). Hereafter, all the relevant historical variables are consequently put on a units per day basis.

As a result, instead of one promotion-response model, we now have eight (four planning periods by fast-versus-slow movers).[18] Our preliminary model results indicated that the 8-model framework was better in terms of average absolute percentage error in units (for a combined 22% improvement over the alternative models being considered at that early stage). While improved fit is desired, we were more motivated by the need to reflect widely differing promotion-duration policies as we deployed this method across grocery chains. Some chains run month-long coupon books, others run major ads for one week and TRP-only as a second week follow-up, and yet others insist on "Front End Cap Displays" changing weekly. The possible endogeneity between duration (i.e., the length of the planned promotion) and promotion policy was one of the factors which persuaded us to set up separate models for the four primary planning periods. The separation into eight models also makes it easier to detect when *not* to make a forecast. For example, if a store's policy is never to have an A Ad for more than three weeks, in the 28-day model, this parameter will be singular. The other parameters will be unaffected, but an

[18]These eight models assume that all relevant historical information is available at the item-by-store-by-promotion-type level. While this is true for over 99% of the promotion events in our samples, additional models were created to handle the different patterns of missing data. The results of these models will not be discussed here to conserve space and due to their limited range of applicability.

automatic warning can be generated not to forecast for this condition, which has never been tried. With a single model, in which a complex pattern of interactions is used to deal with different durations and slow versus standard movers, this condition might be much harder to detect with merely the visual inspection of parameters currently used.

Sampling of Items

When sufficient historical data are available, the population of events from which the sample is drawn consists of all promotions, for any of the UPCs in the item master file, in the most recent 30 months that had durations longer than two days. A minimum of one year of data is needed to calibrate the seasonal variables.[19] A 4 (duration range) by 9 (promotion condition)[20] grid is used to generate a stratified random sample containing approximately 2.5 million promotion events.[21] Note that here we sample promotion events (the basic unit for which we wish a tactical forecast), rather than sample items. This sampling scheme is enabled by having tables that reflect the relevant historical averages for every UPC in the item master file (support tables that were not available during the preliminary development stage). The sample is randomly split, so that half is used for calibration of the models and half is used for cross-checking. The validity results presented below are all conducted on "out of sample" promotion events.

Results

Comparison with Historical Averages on Matching Ad and Display Conditions

The most widespread industry approach to promotional analysis is to compare the average baseline sales for a category (subcommodity) to the *lift* (i.e., the multiple of baseline sales expected when *any* item in a category is promoted in a particular way).[22] Item-to-item sales variation in a category virtually guarantees that using average historical item movement for that particular item in that particular store under those particular promotion conditions has less forecast error than this *base-times-lift* approach. Having extensive UPC-level sales histories in each store makes this gain in accuracy possible. Thus, we compare our forecast accuracy with the historical averages that best match the promotional conditions for each particular item. Note that neither the *base-times-lift* approach nor our approach (using item-specific historical averages) uses price as part of the match. The match is based on ad and feature conditions.

Since ordering is generally done in case units (typically 12 units per case), an absolute yardstick for assessing forecast accuracy is the number of case errors incurred. This provides an operational yardstick that is far more concrete than the error relative to the historical average on matching ad and display conditions or more traditional error measures such as the absolute percentage error in units (APEU). The basic problem with APEU is that it grossly over-weights errors in slow-moving items. A one-unit error is trivial in most instances, but reflects a 100% APEU in an item that sells only one unit. We focus on the operationally relevant error criterion—case errors. If one misses a forecast by a single unit, but that unit error requires ordering another case (resulting in ordering one too many cases), then one has a one case error. If one misses a forecast by 11 units, but still has the correct case order, then no case error is incurred. We will report these values only for the validity studies, since comparisons based on the developmental data may not be sufficiently rigorous.

Longitudinal Cross-Validation

Two cross-validations were undertaken to assess forecast accuracy in contexts closer to real application. The first validation study involved forecasts across trading

[19] While the full time frame was used in pilot market 1, only 15 months of data were available for the second pilot market. The shorter time period had no discernible effect on the statistical properties of the forecasts.

[20] These are the nine matching ad and display conditions defined in the Appendix.

[21] A small (less than 1%) coverage sample is drawn at random, but disproportionately, to ensure as much as possible that all 36 cells in the stratification have events in them.

[22] Earlier we discussed the most sophisticated industry practice (matching on "last-like" promotion price). We report here the most frequently used approach.

areas and over time. We estimated a single set of parameters for three geographically separated trading areas combined (311 stores of a major grocery chain). After calibration (but before the promotions occurred), we forecast the results of 20,710 upcoming events. After the results came in, we found that 69% of the forecasts were within ± one case and 83% of the forecasts were within ± two cases. The corresponding historical averages matched on ad and display conditions predicted only 39% within ± one case and 62% of the events within ± two cases. For both our forecast and the matching historical averages we used the actual store conditions (feature and display) as the basis for our estimates.

The second longitudinal cross-validation is a worst case scenario. It involves the client in pilot market 2 selecting 212 fast-moving items and not being specific about how these items would be promoted in the 178 stores in this client's market area. The ad conditions were generally known, but the in-store display conditions were not known. The chain policy was to do a major display, but whether that display was to be a "Store Front/Promotion Display," "Front End Cap Display," "Rear End Cap Display," or "Free Standing In-Aisle Display" was left to the discretion of the store manager. Whether the store manager actually put up the display was unknown, as our forecast was made after the calibration period, but before the promotion events took place. Display compliance is one of the most uncertain areas of promotion policy. In the first longitudinal cross-validation, *ems, inc.* audited the in-store conditions of display, and what actually occurred was used as the type of display for the forecast. In this second longitudinal cross-validation, we had to guess what might be done for the 30,569 promotion events that these 212 items generated during the three-week test period. We guessed "Rear End Cap Display" as a type of average major display. Still, 66% of the forecasts were within ± one case, and 80% of the forecasts were within ± two cases. Matching historical averages did pretty well, with 63% of the forecasts within ± one case, and 75% of the forecasts within ± two cases.

Discussion

The validation results presented in the previous section are representative of all analyses performed so far. At this writing, we have run two waves of analysis (separated by six months) in pilot market 1, a single analysis of pilot market 2, and preliminary analyses (without validation studies) in four other pilot markets. The preliminary results presented were for the second wave in pilot market 1.

The second validation study provided an opportunity to look more carefully at management decisions concerning ordering. In the Introduction, we indicated that the relative cost of errors was such that it was much better to overstock than to run the risk of out-of-stock conditions. The management for the retail chain in pilot market 2 believed in overstocking. For the 212 fast-moving items the client selected for the pilot study, they stocked 443,000 cases and sold 171,000 (179% safety stock). This is expensive ($26,000 in cost of capital for the three-week pilot). Yet, we estimate they were still out-of-stock on almost $247,000 worth of sales.

We acknowledge that determining this $247,000 figure is problematic. Still, we believe it is possible to derive a reasonable approximation. First, we approximated out-of-stock using a straightforward algorithm on the daily data. We argue that, if an item is selling one or two units per day and sells nothing one day, this might be part of a normal sales pattern. We would not want to claim all zero-movement days as out-of-stock. Only when sales are substantially less than average might we have an out-of-stock condition. But out-of-stock can occur in the middle of a day and grossly distort what the average sales are for the in-stock days. Thus, to be included in the "average" a day's sales had to be no less than 1/3 the maximum sold in any day of the event. For all days in the event that qualified, we computed the average and the standard deviation. If sales on a particular day were more than two standard deviations below "average" sales, we counted the day as having an out-of-stock condition. This algorithm picked up almost all of the zero-movement days in the middle of the sales stream for these fast movers. We conservatively estimated the lost sales due to out-of-stock condition to be 80% of the average sales (minus the actual sales). Of the 170,792 cases sold in the three-week test for these 212 items, we estimated that 19,858 case sales were lost due to being out-of-stock. These lost sales had a retail value of $246,693.

If our forecast had been used (even without a safety stock), 8,514 (43%) of these cases would have been recovered. This is based on 212 items. Using our forecast, when the 212-item volume is projected up to the approximately $35 million worth of promoted sales for the chain in the three-week test in this trading area, and the three weeks are projected to the annual results, the 43% reduction in the 7% out-of-stock (half the industry average) translates into a $2.77 million profit increase across the 178 stores operated by this retailer in this trading area.[23] If our forecasts (with a reasonable safety stock) were used to reduce overstocking as well as out-of-stock, the profitability increase could be substantially larger.

We are trying to implement a production system for tactical forecasting. Expanding to potentially the 3,700 stores tracked by *ems, inc.* by April of 1999,[24] with each store planning promotions for 400 to 1,000 unique UPCs per week and typically receiving forecasts for four potential promotion combinations, we need a system that can scale to 769,600,000 forecasts a year. This necessitates not only a system of computers that is reliable, available, and scalable, but also a philosophy of forecasting that fits the task. We discuss four aspects of such forecasting below.

First, we have to use the data that are best for forecasting and accept the consequences for calibration. For example, in forecasting, we have no part-whole relation between a promotion event being forecast and the historical averages that are an integral part of that forecast. The historical information represents the strategic asset that we desire to exploit. If this means that an event used in calibration is one of the approximately 19 events incorporated into the historic average, so be it. We will accept the part-whole relation in calibration in order to use the best data in forecasting. Using only longitudinal cross-validation to assess the effectiveness of the model eliminates any capitalization on chance that might inflate effectiveness measures from comparisons based on the calibration time period.

Second, we recognize that any method using historical data as a strategic asset has inherent limitations. When a promotion style for a particular item has been used elsewhere in the chain, but never used in the particular store we wish to forecast, our forecast error increases. When an item has previously been promoted in the chain, but never by a particular promotion style that we now wish to forecast, our forecast error increases. When an item has history in some stores in the chain regarding the promotion style of interest, but is new to a store we wish to forecast, our forecast error increases. When an item has history in some stores in the chain regarding other promotion styles but none concerning the promotion style of interest, our forecast error increases. All of these increases in forecast error reflect the cost of using less specific historical data, where better data are not available. Over 99.5% of the time, we have full data from which to create a forecast. However, with a database of 20 million promotion events, 100,000 events have less than ideal data. In calibration, the fit of the models dropped approximately 3% with each of these four degradations in data. When a store has no experience on any item regarding a particular promotion style, we do not attempt to forecast. When an item is new to the chain, we do not attempt to forecast. Other models and methods are more tailored to new product tracking.

Third, any method that wishes to be applicable to every UPC in the item master file has practical limitations. We cannot incorporate competitive information in the forecast. If two major brands are on promotion in the same store, the likely competitive effects will not be taken into account in our forecast. If market-share models could be implemented on this scale, they would undoubtedly do better. But can retailers be expected to develop and maintain market-share models in each of the 1,200 categories they manage in their stores? We have seen no indication of willingness so far. The forecast also does not make use of information on the time since last promotion. This is a limitation of current databases that may be eliminated in the future.

[23]The 7% out-of-stock projects to $2.48 million of lost sales for the retailer during these three weeks. A 43% savings in out-of-stock from our model projects to $160,000 profit at 15% gross margin, which implies $2.77 million on an annual basis for this retailer in this trading area. This translates to $15,500 per store.

[24]This represents a 500-store growth since the September 1997 figures reported at the beginning of this article.

We also recognized early on that the Christmas-New Year period has a special character. Models designed for just this period are on the future agenda.

Finally, if the retail environment is rapidly changing and if retailers or consumers alter their behavior so that the underlying structure changes or conditions change (i.e. are not stationary), we need to recalibrate our model. We have to be able to bring a new chain market online or recalibrate an existing chain market in a week to ten days. Much of that time is spent performing the complex queries that form the support tables for the forecasts, drawn from the event history files for each chain market. This favors the straightforward estimation of parameters in a regression structure over methods that would require much more extensive computer resources to calibrate.

We are indeed operating in a rapidly changing retail environment. Shopper loyalty programs are gaining momentum, and the databases have not yet caught up. Loyal shoppers are offered a variety of discounts that are not yet recorded in the promotion-event databases. New approaches will be needed to reflect the growing reality that multiple discounts are available in a store at any point in time.

By and large, the variables we have incorporated so far have been continuous measures that easily fit into regression-style models. What we have left behind are the categorical measures that deal with the 1,200 subcommodities, or the 1,000 manufacturers. Regression models are not robust to the inclusion of 2,200 additional dummy variables or the possible interactions of such variables with other measures of interest. We are currently working with datamining techniques to find systematic knowledge in such sources and adjoin them to the forecast by using them to explain errors in the current forecast (Cooper and Giuffrida 2000).

We consider PromoCast™ to be a first effort. Just as looking through forecast errors helped improve our preliminary models, we believe that we can learn much more with experience. Hoch and Schkade (1996) report that managers with data and models are much better at forecasting than those with just data or those with just models. We hope that providing data designed to support the promotion planning task and models tailored to the retailer's problems will benefit managers seeking to plan in the complex, modern information environments.[25]

Appendix—Measurement and Definition of Variables in the Model

Ad Measurement

A newspaper advertisement with a large number of items relating to a single grocery retailer is called a feature. *ems, inc.* codes each feature by its appearance to the customer, i.e.,

A - Dominant ad—large ad within a single feature.
B - Secondary ad—medium ad within a single feature.
C - Small/Line ad—small ad within a single feature.
P - Coupon ad—coupon ad within a single feature.

When coding print advertising, *ems, inc.* is concerned with identifying the overall retailer/ad circular (feature) as well as each specific ad feature. Each section or ad is a part that refers to a specific product or set of products. When coding an ad, the first area of concern is whether the ad stands out as a dominant part of the overall feature. The ad or ads that stand out from the rest of the ads are coded as "A" ads. Ad size is the only determinant of ad dominance. Bold print and color should not enter into the coding. When determining size, pictures and/or additional print are included. Occasionally, coders see small ads with small pictures included. These are still considered small.

If a store or manufacturer coupon is included in the feature, it is coded as a "P" ad. Coupon ad coding takes precedence over "A," "B," or "C" ad types. If the ad is a coupon, it is a "P" ad irrespective of whether the product is part of an "A" or "B" ad.

If a coupon ad states "buy one product (x) get a different product (y) free or at another price," the first product (x), the only that the coupon is good for, is coded as a "P" ad. The other product (y) is coded as an "A," "B," or "C," based on the size of the ad. In the interactions listed below, these ad types are collapsed into Major Ads (types "A" and "B") and Minor Ads (types "C" and "P").

Display Measurement

All displays are audited weekly in person in each store by *ems*. The first seven display types are listed below and are self-explanatory. A Secondary Location Display (DISPLAY9) is indicated when, in addition the normal shelf space, a display in a separate part of the store is recorded. Other Display (DISPLAY8) is a rarely used classification for a display that is not properly classified as DISPLAY1-DISPLAY7 or DISPLAY9. In the interactions listed below, these display types are collapsed into Major Displays ("Store Front/

[25]The authors had the support and assistance of many individuals in undertaking this project. While they are indebted to them all, they especially wish to thank Giovanni Giuffrida (doctoral candidate in Computer Science at UCLA), and Bill Weissenberg and Brian Rock (*ems, inc.*) for their outstanding efforts on this project. They gratefully acknowledge the helpful comments of Don Morrison. They also thank Kimberly Weissenberg and Sharon Bear, Ph.D. (BearWrite @aol.com) for their editorial assistance.

Promotion Display," "Front End Cap Display," and "Rear End Cap Display") and Minor Displays ("Store Rear Display," "Mid-Aisle End Cap Display," and "Side Aisle End Cap Display").

Best Matching Historical Averages

When we refer to "matching" or "best-matching" promotions, only ad and display conditions are considered. Price is ignored since it has too many levels to be a basis of a practical-matching scheme. Further, in this version of our model, we match only on Major Ad, Minor Ad, or No Ad crossed with Major Display, Minor Display, or No Display. This produces nine discrete conditions for which historical average (daily) sales are computed and stored for each UPC. Each UPC has a store-specific average and a chain-wide average for each of these nine conditions. The same approach is taken to reflecting a store's performance in "best-matching" conditions. For each of these nine conditions, each store's average (daily) unit sales across all UPCs is stored in support tables.

Variable	Variable Definition
LNUNITS	The dependent variable is measured as the log of total units sold during the promotion event. This measure is specific to each UPC and to each store.
INTERCEP	The standard intercept in a regression model.
UNITPR	Price per unit sold.
PO	Let S be the shelf price and U be the unit price during the promotion event. Then PO is $1/(1 + [(S - U)/S])$.
X_FER	An indicator variable [0,1] that takes the value 1 whenever X units are offered for the price of Y units (e.g. "2-for-1" sales).
ACVDOLS	The total dollars spent for all purchases in the subcommodity during the calibration period.
LACVDOLS	The log of ACVDOLS.
A_AD	An indicator variable [0,1] that takes the value 1 whenever an A Ad is present.
B_AD	An indicator variable [0,1] that takes the value 1 whenever a B Ad is present.
C_AD	An indicator variable [0,1] that takes the value 1 whenever a C Ad is present.
P_AD	An indicator variable [0,1] that takes the value 1 whenever a P Ad is present.
DISPLAY1	An indicator variable [0,1] that takes the value 1 whenever a Store Front/Promotion Display is present.
DISPLAY2	An indicator variable [0,1] that takes the value 1 whenever a Store Rear Display is present.
DISPLAY3	An indicator variable [0,1] that takes the value 1 whenever a Front End Cap Display is present.
DISPLAY4	An indicator variable [0,1] that takes the value 1 whenever a Mid-Aisle End Cap Display is present.
DISPLAY5	An indicator variable [0,1] that takes the value 1 whenever a Rear End Cap Display is present.
DISPLAY6	An indicator variable [0,1] that takes the value 1 whenever a Side Aisle End Cap Display is present.
DISPLAY7	An indicator variable [0,1] that takes the value 1 whenever an In-Aisle Display is present.
DISPLAY8	An indicator variable [0,1] that takes the value 1 whenever an Other Display is present.
DISPLAY9	An indicator variable [,1] that takes the value 1 whenever a Secondary Location Display is present.
MAJAXPO	The two-way interaction of a Major Ad and PO.
MINAXPO	The two-way interaction of a Minor Ad and PO.
MAJDXPO	The two-way interaction of a Major Display and PO.
MINDXPO	The two-way interaction of a Minor Display and PO.
ADJXDISJ	The two-way interaction of a Major Ad and a Major Display.
ADJXDISN	The two-way interaction of a Major Ad and a Minor Display.
ADNXDISJ	The two-way interaction of a Minor Ad and a Major Display.
ADNXDISN	The three-way interaction of a Minor Ad and a Minor Display.
AJXDJPO	The three-way interaction of a Major Ad, a Major Display, and PO.
AJXDNPO	The three-way interaction of a Major Ad, a Minor Display and PO.
ANXDJPO	The three-way interaction of a Minor Ad, a Major Display, and PO.
ANXDNPO	The three-way interaction of a Minor Ad, a Minor Display, and PO.
KEYPAYWK	An indicator variable [0,1] that takes a value of 1 whenever the promotion period contains the 1st or 15th of the month (the Key Pay dates).
SUPERBWL	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to the Super Bowl.
EASTER	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to Easter
MEMORIAL	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to Memorial Day.
FOURTH	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to 4th of July.
LABOR	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to Labor Day.

NEWYEARS	An indicator variable [0,1] that takes a value of 1 whenever the promotion period includes any day in the week leading up to New Years Eve.
INUMAT	Average units per day on matching ad and display conditions for this UPC in this store during the calibration period.
LINUMAT	The log of INUMAT.
INBMAT	The record of promotion event includes the baseline sales for this UPC in this store at the time of the promotion event. INBMAT is the average of these baseline sales for this UPC in this store on matching ad and display conditions during the calibration period.
LINBMAT	The log of INBMAT.
INHU	Average units per day on across all promotions for this UPC in this store during the calibration period.
LINHU	The log of INHU.
INHB	The average of the baseline sales for this UPC in this store across all promotions during the calibration period.
LINHB	The log of INHB.
INHN	Number of promotion events for this UPC in this store during the calibration period.
IUMATCH	Average units per day on matching ad and display conditions for this UPC across all stores during the calibration period.
LIUMATCH	The log of IUMATCH.
IBMATCH	The average of the baseline sales for this UPC across all stores on matching ad and display conditions during the calibration period.
LIBMATCH	The log of IBMATCH.
IHU	Average units per day across all promotions for this UPC across all stores during the calibration period.
IHB	Average baseline sales across all promotions for this UPC across all stores during the calibration period.
NUMATCH	Average units per day on matching ad and display conditions across all UPCs in this store during the calibration period.
NBMATCH	Average baseline sales on matching ad and display conditions across all UPCs in this store during the calibration period.
POXINU	The two-way interaction of PO and INUMAT.
ISB	An indicator variable [0,1] that takes a value of 1 whenever the matching promotion style for this UPC in this store is much better than the average across all promotion styles for this UPC in this store (i.e., when 0.6*INUMAT > INHU) during the calibration period.
ISB1	The two-way interaction of ISB with INUMAT.
ISB2	The two-way interaction of ISB with INHU.
ISW	An indicator variable [0,1] that takes a value of 1 whenever the matching promotion style for this UPC in this store is much worse than the average across all promotion styles for this UPC in this store (i.e., when INUMAT < 0.6*INHU) during the calibration period.
ISW1	The two-way interaction of ISW with INUMAT.
ISW2	The two-way interaction of ISW with INHU.
SGTC	An indicator variable [0,1] that takes a value of 1 whenever the matching promotion style for this UPC in this store is much better than the matching promotion style for this UPC across all stores (i.e., when 0.6*INUMAT > IHU) during the calibration period.
SGTCX1	The two-way interaction of SGTC with INUMAT.
SGTCX2	The two-way interaction of SGTC with IUMATCH.
SLTC	An indicator variable [0,1] that takes a value of 1 whenever the matching promotion style for this UPC in this store is much worse than the matching promotion style for this UPC across all stores (i.e., when INUMAT < 0.6*IHU) during the calibration period.
SLTCX1	The two-way interaction of SLTC with INUMAT.
SLTCX2	The two-way interaction of SLTC with IUMATCH.

References

Belsley, David A., Edwin Kuh, Roy E. Welsh. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley & Sons, New York.

Blattberg, Robert C., Scott A. Neslin. 1990. *Sales Promotion: Concepts, Methods and Strategies.* Prentice-Hall, Englewood Cliffs, NJ.

Bridges, Eileen, Chi Kin (Bennett) Yin, Richard A. Briesch. 1995. A high-tech product market share model with customer expectations. *Marketing Sci.* **14** (1 Winter) 61–81.

Christen, Marcus, Sachin Gupta, John C. Porter, Richard Staelin, Dick R. Wittink. 1997. Using market-level data to understand promotion effects in a nonlinear model. *J. Marketing Res.* **34** (August) 322–334.

Cooper, Lee G., Giovanni Giuffrida. 2000. Turning datamining into a management science tool: new algorithms and empirical results. *Management Sci.* (forthcoming).

——, Masao Nakanishi. 1988. *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness.* Kluwer Academic Publishers, Boston, MA.

DeKimpe, Marnik G., Dominique M. Hanssens. 1995. Empirical generalizations about market evolution and stationarity. *Marketing Sci.* **14** (3, Part 2) 109–121.

Finney, D. J. 1941. On the distribution of a variate whose logarithm is normally distributed. *Supplement to J. Royal Statist. Assoc.* **7** (2) (issued with Vol. 104, Part IV of the journal), 155–161.

Franses, Philip Hans. 1995. The effects of seasonally adjusting a periodic autoregressive process. *Comput. Statist. and Data Anal.* **19** 683–704.

——, Richard Paap. 1995. Moving average filters and periodic integration. *Math. and Computers in Simulation* **39** 245–250.

——, Svend Hylleberg, Hahn S. Lee 1995. Spurious deterministic seasonality. *Econom. Letters* **48** 241–248.

Hoch, Stephen J., David A. Schkade. 1996. A psychological approach to decision support systems. *Management Sci.* **42** (1) 51–64.

Leone, Robert P. 1995. How promotions work. *Marketing Sci.* **14** (3, Part 2) 141–150.

Lodish, Leonard M., Magid M. Abraham, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, Mary Ellen Stevens. 1995. A summary of fifty-five in-market experimental estimates of the long-term effect of TV advertising. *Marketing Sci.* **14** (3, Part 2) 133–140.

Miller, Don M. 1984. Reducing transformation bias in curve fitting. *Am. Statistician* **38** (2) 124–126.

Neyman, Jerzy, Elizabeth L. Scott. 1960. Correction for bias introduced by a transformation of variables. *Ann. Math. Statist.* **31** (Part 2, Sept.–Dec.) 643–655.

Wittink, Dick R., Michael J. Addona, William J. Hawkes, John C. Porter. 1987. SCAN*PRO®: a model to measure short-term effects of promotional activities on brand sales, based on store-level scanner data. Working Paper, Johnson Graduate School of Management, Cornell University, Ithaca, NY (May).

This paper was received December 3, 1997, and has been with the authors 6 months for 3 revisions; processed by Richard Staelin.