
Does TV Advertising Really Affect Sales? The Role of Measures, Models, and Data Aggregation

Gerard J. Tellis and Doyle L. Weiss

Traditional econometric models suggest that advertising has a clear and significant positive effect on sales in the current period. However, recent studies using scanner data indicate that the estimated effects of TV advertising on households' brand choices are weak and rarely significant. Do those findings mean that TV advertising does not really have an impact on current brand choices and sales? Or are the discrepant findings due to differences among the measures, models, and aggregation levels used by different researchers? The authors address these issues. Their analysis indicates that aggregating data over time and households may create a false impression of advertising having a statistically significant effect on sales.

Gerard J. Tellis is the Jerry and Nancy Neely Professor of Marketing, University of Southern California, Los Angeles.

Doyle L. Weiss is Murray Professor of Marketing, University of Iowa, Iowa City.

The authors thank the Advertising Research Foundation and Information Resources Inc. for the data, Sharat Mathur for his help in the analysis, and Avu Sankaralingam for his comments. The article also benefited from the comments of the editor and anonymous reviewers. The study was partly supported by a Marketing Science Institute grant to the first author.

Marketers continue to spend large amounts of money on advertising, which is one of the most important and visible marketing tools. Advertising is generally thought to have a current-period influence on sales, called the current effect, and a long-term influence beyond the current period, called the long-term or carryover effect. For decades, researchers have used econometric models to estimate both of these effects of advertising to help firms set their advertising budgets. Our review of these models has identified two distinct streams of research, which we label sales response modeling and choice modeling. The two research streams are similar in goals and orientation, but differ substantially in their data and models. The results of studies on the effects of advertising from these two streams of research are also substantially different. We first describe the differences in data, models, and results from these two streams of research, and then describe the goals of the study we conducted to determine a cause for the difference.

Sales Response Models

One stream of research uses sales response models to find the responsiveness or elasticity of sales to current-period advertising while controlling for other variables, typically price. These models use data that generally are aggregated over all buyers, segments, and regions of a country to the national level. Temporally, the models aggregate the data into annual, quarterly, monthly, or weekly periods. Thus, sales response models use data that are aggregated both temporally over purchase occasions and cross-sectionally over buyers. The models typically use sales or market share as the dependent variable, and analyze the effects of advertising, price, and other variables on these dependent variables.

The large number of studies using sales response models yield some broad generalizations about the effects of advertising on sales or market share. These effects are measured primarily as the sales response or the sales elasticity to advertising. On the basis of such models for numerous European packaged goods brands, Lambin (1976) found the mean of 40 significant advertising elasticities to be .10. Leone and Schultz's (1980) early review of sales response models found advertising elasticities to range from .003 to .482, most being below .2. In a meta-analysis of these models,

Assmus, Farley and Lehmann (1984) concluded that the elasticity of sales to advertising was .22. A more extensive meta-analysis by Sethuraman and Tellis (1991) found that the elasticity of sales to advertising was .1. This stream of research suggests that the effects of advertising on sales are both measurable and significant.

Choice Models

Recently, another research stream has emerged in which choice models analyzing single-source data are used to address the same issues. Single-source data are a combination of scanner data and cable data (Curry 1993). Scanner data are records of consumer purchases collected by laser scanners at the checkout point of grocery stores. These data are very disaggregate, containing detailed records of individual households' purchases by item, promotional status, and minute of occurrence. Promotional status includes the price, the presence of in-store displays, retail advertising measures, and records of coupons redeemed at the time of purchase. Cable data are records of the same households' exposures to TV advertising at five-second intervals. The combination of scanner data and cable data are called single-source data because it provides in one place the key information that affects consumers' purchases of grocery products. The strength of the data is in their wide coverage of key variables and their in-depth coverage of purchase timing and choices of individual items by individual households. (For more information on the collection and uses of these data, see Tellis 1988a; Assael and Poltrack 1992; and especially Curry 1993).

Choice models (Guadagni and Little 1983; McFadden 1974) are a powerful method for analyzing single-source data, because they do so at the highly disaggregate level at which consumer decisions are made. In particular, choice models can analyze individual household choices in terms of brands, time, and quantity purchased as a function of advertising and promotional variables. Whereas the earlier stream of research on advertising effects used sales response models, the newer stream uses choice models. Comparing choice models with sales response models in econometric research is analogous to comparing the electron microscope with the optical microscope in biological research or the laser with the scalpel in surgery. The new data and models are better not merely because they are disaggregate, but because they enable us to investigate phenomena at the disaggregate level at which the phenomena occur. The

new data and models promise great insights into consumer behavior and the potential elimination of bias in estimates of consumer response to advertising and promotion. Several authors consider the new data and models to be no less than a revolution in marketing research (e.g., Assael and Poltrack 1992; Curry 1993; Guadagni and Little 1983).

A growing consensus among researchers working with choice models is that the current effects of TV ad exposures on purchases are too small to capture or perhaps nonexistent. In contrast, the current effects of price, in-store displays, and retail newspaper ads are pronounced and difficult to miss. For example, in a choice analysis of toilet tissue purchases, Tellis (1988) found the main effects of TV advertising to be marginal. In addition, the Advertising Research Foundation initiated five studies in which single-source data were used to determine the effects of advertising. Only one of the studies found any effects worth reporting. That study (Kanetkar, Weinberg, and Weiss 1989) found no main effect, but instead a negative interaction effect of advertising on price elasticity. The effect implies that advertising increases price sensitivity and plays a role opposite to that intended by managers. In the only published choice model of which we are aware that finds significant positive effects of advertising using single-source data, the effect sizes are small (Pedrick and Zufryden 1992). A few unpublished studies may show positive significant effects, but probably many more show no significant effects. Indeed, many researchers who analyze scanner data with choice models do not include advertising in their models (e.g., Guadagni and Little 1983; Gupta 1988; Lattin and Bucklin 1990).

Scope of Current Study

Sales response models and choice models lead to different findings about the estimated effects of TV advertising. Are the discrepant findings due to differences in time periods, advertising media or product categories? Or are they due to differences in models, measures of advertising, or data aggregation? If one of the latter is the reason, what is the appropriate model, measure, or level of aggregation for estimating advertising's real effects?

These questions motivated our study. In particular, it had the following objectives:

- To determine whether the current effects of TV advertising vary by temporal or cross-sectional aggregation.

- To determine whether the current effects of TV advertising vary by some alternate measures and models used in the literature.

To achieve these objectives, we analyzed one category, laundry detergents. We used detergent data for two reasons. First, the heavy use of TV advertising for the category suggests that its marketers believe such advertising works. Also, because of the heavy advertising in the category, the probability of our identifying some of advertising's effects was higher. Second, in the category advertising patterns vary greatly across brands and over time. Higher variance in an independent variable increases the probability of correctly estimating the effects of that variable.

We evaluated the variance in estimates of TV advertising's effects on household purchases in this category over different models, measures of advertising, and aggregation levels. We contrasted the estimates from choice models with those from sales response models. For each model, we contrasted the estimates from different measures of advertising. For the regression models, we contrasted the estimates from data aggregated at the household level with those from data aggregated at the market level comprising of the total of all households; we also contrasted the estimates from data aggregated in days, weeks, and months.

Our approach was primarily empirical because the precision of data, measures, and estimation methods greatly exceeds the precision of the theories about advertising's effects. Our analysis was not meant to deny the value of theory but to augment research into theory.

The Problem of Data Aggregation

In the past, most data were aggregated across individuals and across time periods. The question of aggregation level was moot. However, with scanner panel data available at the level of household choice, the selection of an operational aggregation level has become important. Our purpose was to study advertising's effects on sales, which is a sum of individual consumer choices. Advertising only can affect sales through the effect of ad exposures on these choices of individuals. Therefore, for our purposes, we believed that data at the level of individual households, reflecting advertising exposures and individual purchases, was likely to produce the most reliable estimates of advertising's effects. We discuss this issue in more detail subsequently. Here we briefly re-

view the history of the estimation of the effects of aggregation on advertising response.

In econometrics, disaggregate data are generally preferable to aggregate data. The reason is that data aggregation always leads to a loss of statistical power and efficiency (Judge et al. 1985). In addition, data aggregation can bias the estimated effects of an independent variable on the dependent variable (Cooper and Nakanishi 1988; Judge et al. 1985). One aggregation issue covered extensively in the marketing literature is of the carryover effects of advertising. Such effects may be due to delays in purchase after exposure to an ad, long-term memory for the ad, word-of-mouth diffusion of the ad's message, or other factors.

Beginning with a pioneering study by Palda (1964), researchers have tried to measure the carryover effects of advertising, typically with one class of models called Koyck models. The appeal of these models is that, subject to some assumptions, they capture all prior or carryover effects of past advertising through the coefficient of one variable. That variable is the lagged value of the dependent variable that is included as an independent variable in the model. In an important review in this tradition, Clarke (1976) argued that data aggregation upwardly biases the estimates of the carryover effects of advertising. Subsequently, researchers have shown the cause of the bias, estimated its size, and provided models to correct for it. Only a few studies have examined the data aggregation in estimates of the current effects of advertising. Clarke (1976), Windal and Weiss (1983), and Bass and Leone (1983, 1986) showed or noted that, if unaccounted for by the specification, the estimate of advertising's current effects will increase as the data aggregation increases.

Our study differed from the studies cited in one important respect: we examined the role of bias, if any, in the estimate of the current effects of TV advertising due to measures, models, or data aggregation. We examined effects of both cross-sectional and temporal aggregation. Most of the studies on carryover effects were fairly narrow in scope. They (1) were restricted to sales response models, (2) often assumed that the true response to advertising followed a Koyck process, (3) used data that were aggregated over all buyers, and (4) measured advertising in dollars or gross ratings points. With today's richer single-source data, we can measure sales and advertising exposure very accurately for individual households and brands. Analyses of such data using choice models generally fail to find significant estimates of any current effects of TV advertising on choices. Sales response models,

in contrast, usually show a significant positive effect of such advertising on sales. We therefore addressed the question of whether the findings from these two research streams differ due to models, measures or data aggregation?

Data

We used single-source data collected by IRI (Information Resources Incorporated) and given to us through the Advertising Research Foundation as part of a major study initiated by the latter organization. Five categories of data were made available, from which we chose laundry detergent because it had the highest and most intensive level of TV advertising. Based on figures from Leading National Advertisers, we estimated that all the brands in the category spent about \$260 million on TV advertising for the year we studied. Such heavy advertising implies that firms believed in the merits of TV advertising, and therefore such advertising should have measurable effects on consumer choices. Of the 28 brands in the category, 16 advertised. Brands owned by Procter and Gamble dominated the category, capturing over 50% of the purchases and TV exposures recorded. For the entire category, the records covered 88 weeks of household purchases of laundry detergents for 1983-1984 in Eau Claire, Wisconsin.

Panelists' Purchases

We used the first 30 weeks of data to initialize some variables, and the remaining 58 weeks of data to estimate the models. The original sample consisted of about 2000 panelists, 970 of whom had meters that recorded their TV viewing. We dropped panelists who had purchased less than three times in either the calibration or the estimation period, and panelists whose TV meter worked less than 81% of the time, which left 162 panelists. The sample was somewhat small, but previous studies used as few as 100 panelists (e.g., Guadagni and Little 1983).

Choice Sets

We defined the purchase alternatives available in the laundry detergent category for four levels of detail: brands, forms, sizes, and UPCs. We did not drop any brands to preserve the string of consecutive purchases by households. In particular, we used 28 brands, two forms (powder, liquid) for each brand, many sizes for each form, and a few different UPCs

(the item identification) for each size. Because pricing and promotion tend to be specific to a brand, form, and size, we did not aggregate over any of those three levels. The only aggregation (in the data processing) was across UPCs with the same brand, size, and form. Over the whole market, and time period, this process yielded 93 alternatives available in the market. An alternative was a unique brand, size, and form available in a few UPCs (e.g., Tide 32 oz. liquid).

In merging with the purchase data, we reduced this number (93) to a more plausible choice set. First, the choice set was defined to be the alternatives available within the store in which the panelist buys. Second, only alternatives bought at least once (by some panelist) in a four-week period were assumed to be available in the store. We refer to this four-week period as a calibration month. Third, we classified these alternatives into 11 categories (choice sets) based on form and size, five liquid and six powder. For each occasion and store visit, we assigned panelists to one category according to the alternative chosen. Our assumption was that consumers are not likely to be in the market for alternatives of widely different sizes or forms. Hence, our analysis of choices and sales was conditional on purchase occasion, store, and choice set decisions (which were not explicitly modeled). These limitations are not critical because our focus was TV advertising, which is not likely to affect decisions of timing, store, or choice set. Finally, our comparison was across different models, measures, and aggregation levels, to all of which the same data preparation applied.

Marketing Variables

We obtained information on the prices, displays, features, and coupons from the purchase data and the store data. Prices were the actual prices paid by panelists for a particular brand, in a particular store, at the minute of purchase. A typical aspect of such prices is that they fluctuate from week to week as the retailer offers discounts, but remain the same within a week. However, the price we used was not the weekly stable list price, but a transaction price to which the value of coupons (if used) was added back. The recorded transaction price was net of coupons. So, for the brand a consumer bought, we defined the list price as the transaction price plus the value of any coupons used. For a brand a consumer did not buy, we defined the list price in a store and week as the highest modal list price paid for that brand by any panelist in that store for that week. For weeks in

which no purchase occurred at all, the list price was that of the previous week in the calibration month.

In-store displays are physical setups sent or paid for by the manufacturer to provide greater visibility for the brand. Not all retailers accept displays or have them during the same week. Hence, displays vary by store besides varying by brand and week. Past research, as well as our perusal of the data, suggested that displays substantially increase sales. We therefore included a dummy variable for the presence or absence of displays for all panelists shopping for a brand in the store that had the display during a particular week.

Features are advertisements by the retailer in the local media announcing the availability or price of a brand. Most features are free-standing inserts in the local newspapers. Some features may be mailed or distributed to households. Features generally indicate that a brand has a special lower price during a particular week in the store, although some may only indicate availability. We included a dummy variable for the presence or absence of features for all panelists shopping for a brand in the store that had the feature during a particular week. Features may appear collinear with price discounts in aggregate data, but with disaggregate data there is sufficient variation among consumers who do and do not buy on sale, and stores that do and do not feature discounted prices for one or another brand, that collinearity is not a problem.

Coupons are certificates from the manufacturer that guarantee consumers a refund on some portion of the list price at the point of purchase if they meet certain conditions. Because the data did not record the availability of coupons to individual households, we inferred it from the data. We defined coupons as a dummy variable that takes the value 1 for a brand in a week if at least one panelist bought the alternative with a coupon in that week; otherwise it takes the value 0. Note that displays, feature, and coupons measured the availability and not the use of those marketing variables.

TV Ad Exposures

A TV meter records TV ad exposures by brand, minute of day, and five-second intervals. Because the TV meter sometimes breaks down, a separate variable records the days when the meter is not working for each panelist. We dropped panelists whose meter worked less than 81% of the time. The 81% criterion appeared to be the best compromise between having

too small a sample size and having too many missing parts of observations. The meters of the remaining panelists on average worked more than 99% of the time.

Many of our data-processing steps were similar in spirit to those of other researchers (Gupta 1988; Lattin and Bucklin 1989; Tellis 1988a). They were not critical to our study because our purpose was to compare across models, measures, and aggregation levels for which the same data processing applied. Also, we had previously carried out a similar analysis with a fairly different aggregation approach (all alternatives were aggregated to the brand level). The coefficients of price in that analysis were a little weaker than those described here, but the coefficients of advertising were similar in the two analyses.

Brand Loyalty

To control for heterogeneity, we included a simple measure of past behavior, sometimes referred to as brand loyalty. It is defined as the share of choices of each household of each brand in a prior holdout period (e.g., Lattin and Bucklin 1989).

$$(1) \quad L_{ij} = C_{ij}/N_i$$

where

L_{ij} = panelist i 's loyalty for brand j

C_{ij} = panelist i 's choices of j in the prior holdout period

N_i = panelist i 's total choices of all brands in the prior holdout period.

Note that at the individual level, brand loyalty is the preference of one household for one brand. At the aggregate level, it is the average preference of a group of individuals (or segment) for the brand.

Research Design

TV Advertising Measures

All of our measures of advertising were based on recorded TV exposures. Because we had no record of actual consumer viewing, these exposures were strictly only opportunities to see the ad. However, exposures are important because advertising managers can control only exposures, not opportunities to see.

Advertising theory does not provide clear directions for precise measures of advertising, such as incidence or duration of ads, counted independently or in relation to competition, occurring in the current period only or in the last few periods. In most cases we therefore used multiple measures, rerunning a model

for each measure. Because exposures can be computed in several ways, and because the effects of TV advertising may be weak and could be affected by the measures, testing multiple measures of advertising is important.

Households may be influenced by either the incidence of each ad or the duration of each ad, especially if they tune in to only a fraction of the ad. We therefore used both measures of TV exposures (i.e., number of ad exposures and the total duration of ad exposures in a particular time interval). Households may be influenced by a brand's exposure per se or by the impact of that brand's exposure in relation to exposures of other brands in the category. Hence, we used both the independent exposures and the share of exposures in a time interval. The latter variable is often referred to as share-of-voice. Note that we can measure share-of-voice at the aggregate level for the market and at the disaggregate level for each panelist.

Households may be influenced by the exposure in the current time interval only or by past exposures if advertising's effects last beyond the current interval. We included both types of measures. For past advertising, we used a simple lagged value of the last period's advertising as well as an exponentially weighted average of past exposures to compute the stream of past advertising. The latter variable is defined as

$$(2) \quad AS_t = \alpha A_t + (1 - \alpha)AS_{t-1},$$

where A is ad exposure, AS is advertising stock, α is a constant reflecting the decay of advertising's effects, and t is a subscript for time interval. We used the first 30 weeks of data to initialize this moving average. Because theory does not give a clear value of α , we tried different measures between zero and one. An advantage of this single measure of advertising stock is that it avoids multicollinearity from including multiple lags of advertising in the same equation. In the following discussion, advertising refers to TV advertising only.

We crossed all three ways of measuring advertising so that in total we tested eight different measures of advertising: (incidence or duration) \times (independent or share) \times (simple lagged or advertising stock).

Data Aggregation

We explored the influence of two types of data aggregation, cross-sectional and temporal.

Cross-Sectional Aggregation. We carried out the

analysis with data at the individual-household level and aggregated across all households representing a market. In the former case the dependent variable was brand choice (0 or 1) and in the latter case it was sales. This comparison represents the difference between sales response models typical of the earlier stream of research and brand choice models typical of more recent research.

Temporal Aggregation. With scanner data, temporal levels of aggregation can range from seconds to years. We chose days, weeks, and months. (Analysis at the level of seconds, minutes, and hours expands the data to the point that they become almost unmanageable). Because we had only 58 weeks of data, disaggregation and aggregation yielded 406 days of daily data, 58 weeks of weekly data, and 14 periods of monthly data that we could use for our analysis. We did not aggregate the data above the level of months because we would have lost most of the temporal variation, and the cross-sectional influence of interbrand competition would have dominated the data. We believed the comparison across aggregates of days, weeks, and months was adequate for our purpose.

Models

Our primary purpose was to determine how estimates of advertising response from two commonly used models (the logit and regression models) vary across measures and aggregation levels. Because we were already testing several measures and aggregation levels, we restricted our analysis to one simple form of each model. At the individual level, we compared the estimated effects from the logit model with those from the regression model (see Tables 1 and 2). For the temporal and cross-sectional aggregates *beyond the individual level* we used only estimates from the regression model because the logit model can be estimated only on choices occurring at disaggregate purchase occasions.

To analyze sales, we used the following regression model (constant terms omitted for ease of presentation):

$$(3) \quad S_{ijt} = \sum_k \beta_k X_{kijt} + u_{ijt}$$

where S_{ijt} is sales of brand j to panelist i in time period t , X_{kijt} are independent variables describing panelist and brand characteristics of the sales (the variables are brand loyalty, advertising, price, feature, display, and coupons), β_k is a set of parameters

to be estimated, and u_{ijt} are errors assumed to be i.i.d. normal. For ease of exposition in the following discussion, we do not use different superscripts for estimates of β obtained from alternate models.

To analyze sales at aggregate levels we used corresponding aggregates of equation (3).

$$(3a) \quad S_{ijt'} = \sum_k \beta_k X_{kijt'} + u_{ijt'}$$

where the subscript t' implies an aggregate of values of variables in the corresponding t microperiods. Thus, when t is an index of the first seven days in March, t' is an index for the first week of March.

$$(3b) \quad S_{jt} = \sum_k \beta_k X_{kjt} + u_{jt} \text{ for aggregation across households}$$

$$(3c) \quad S_{jt} = \sum_k \beta_k X_{kjt} + u_{jt} \text{ for aggregation across households and time periods}$$

To analyze brand choices, we used the logit model defined as

$$(4) \quad P_{ijt} = 1 / \{1 + \exp(-\sum_k \beta_k X_{kijt})\}$$

where P_{ijt} is the probability that subject i chooses brand j in time period t , subject to errors (in specification and estimation) that are identically and independently distributed Gumbel or type 1 extreme value (Ben-Akiva and Lerman 1989). We used this binary logit formulation because it is comparable to the regression model.

The key difference between a binary logit model and a regression model is that the former assumes an S-shaped response and the latter assumes a linear response. Advertising response is often assumed to be S-shaped because if there are many exposures they may be in the region of saturation, and if there are very few exposures they may be lost in the noise. The logit model is best for capturing an S-shaped response if it is present. An extensive review found no evidence of the S-shaped curve (Simon and Arndt 1980).

Our design ensured that we could separate the influence of models, measures, and aggregation levels.

Results

We report the results separately for the advertising measures, the models, and each type of aggregation. None of the independent variables is standardized, so

the coefficients can be interpreted directly in terms of units of the independent variables. In comparing two estimated coefficients of advertising, if both are not significantly different from zero, we conclude that the two approaches that yield the coefficients are indistinguishable. If one is significantly different from zero and one is not, we conclude the two approaches that yield the coefficients are different. If both coefficients are significantly different from zero, we conclude the two approaches are different only if the coefficients are significantly different from each other as determined by the F-test.

Advertising Measures

We found no differences in estimated coefficients for any of the advertising measures. First, we found no substantial difference in results when advertising was measured independently as the level of a brand's advertising and when it was measured in relation to competition as share-of-voice. Therefore, in subsequent analyses, for ad-stock we used the independent measures of advertising because they are easier to compute and interpret; in all other cases we used share-of-voice because it is managerially more meaningful.

Second, in the regression and logit analyses, we found no difference in estimated coefficients of the two measures of advertising, number of exposures and time of exposure in seconds (see Table 1). For the subsequent analyses we present results only for the number of exposures because it is managerially more meaningful. (Note that Table 2 reports estimates of all independent variables for the logit and regression models, whereas Table 1 gives only estimates of different measures of advertising for those models.)

Third, the estimated coefficients of advertising were not different when advertising was included as ad-stock or as current advertising. We calculated ad-stock by using equation 1, for α values ranging by .10 from zero to one. Including both current advertising and ad-stock in the same model would not make sense (because ad-stock already contains current advertising); also, $\alpha = 1$ implies only current effects of advertising. No value of α yields significant coefficients of advertising. In addition, the estimated coefficients of advertising do not change substantially as α increases. Table 1 gives results for three different values of α .

Models

Table 1 shows that the coefficients of advertising

Table 1
Impact of Different Ad Measures and α Weights on Coefficients of Advertising
in Logit and Regression Models*

A. Logit Estimates of Effect of Advertising on Choice						
Advertising Measure	$\alpha = .8$		$\alpha = .5$		$\alpha = .2$	
	β	χ^2	β	χ^2	β	χ^2
Ad-stock (seconds)	-.001	.14	.002	.34	.001	.15
Ad-stock (exposures)	.017	.02	.11	.77	.08	.47

B. Regression Estimates of Choices on Advertising						
Advertising Measure	$\alpha = .8$		$\alpha = .5$		$\alpha = .2$	
	β	t	β	t	β	t
Ad-stock (duration)	-.0003	1.28	-.0001	.64	-.0002	1.18
Ad-stock (incidence)	-.007	.66	-.002	.26	-.006	.78

* Each row represents results from a run with a different model or measure of advertising. Effects of other independent variables are not reported here for simplicity. See Table 2 for the latter estimates.

are insensitive to type of model. As long as the analysis is at the individual-panelist level and does not involve cross-sectional or temporal aggregation, the coefficients of advertising are not significantly different from zero. Table 2 compares the coefficients of all independent variables estimated by the logit and regression models. Although the coefficients are not directly comparable, the χ^2 and t statistics may be compared. The χ^2 is equivalent to the square of the t-statistic. Note that the rank orders of these statistics are similar in the two models. Also, in both models, all variables but advertising have coefficients that are significantly different from zero. Finally, consistent with findings in other studies, the test statistics (χ^2 or t) for loyalty are substantially higher than those for other variables.

Data Aggregation

Temporal Aggregation. Table 3 shows the influence of temporal aggregation. Note that the coefficients of most of the independent variables do not change much with temporal aggregation, whereas those of advertising increase systematically with temporal aggregation. For example, the coefficient of lagged share-of-voice increases from -.003 (t = .4) at the daily level,

to .02 (t = 1.7) at the weekly level, to .07 (t = 2.7) at the monthly level. This finding is generally consistent with behavior noted by Clarke (1976), Bass and Leone (1983,1986), and Windal and Weiss (1980).

Cross-Sectional Aggregation. Table 4 reports the influence of cross-sectional and temporal aggregation. The market level is a (cross-sectional) aggregation of all individual panelists' observations (equation 2b). To make a simple comparison, Table 4 gives the influence of aggregation on the estimates of coefficients for only two variables, the lagged share-of-voice and coupons. Note that the coefficients of advertising increase with temporal aggregation at both the individual-panelist level and of the market-aggregate level. Also, coefficients of lagged share-of-voice are higher at the market-aggregate level than at the panelist level for every level of temporal aggregation. In contrast, the coefficients of the other independent variables do not change much, as Table 3 also shows. In particular, Table 4B shows the insensitivity of the coupon coefficients to both temporal and cross-sectional aggregation.

Discussion

The key conclusions from our analyses are that the

Table 2
Comparison of Estimates by Logit and Regression Models
(Dependent Variable: Brand Choice)

Independent Variable	<u>Logit Estimates</u>		<u>Regression Estimates</u>	
	β	χ^2	β	t
Cumulative exposure	.11	.8	-.002	.3
Loyalty	5.87	1487.9	.87	69.1
Price	-.63	12.4	-.02	2.7
Coupon	1.26	295.2	.10	16.8
Display	1.87	246.6	.27	21.3
Feature	.72	11.5	.13	5.8
"R ² ", R ²	.36		.36	

estimated effects of advertising, especially the current effects:

- are not very sensitive to alternate measures,
- are not very sensitive to alternate models (e.g., regression or logit) if both are at the same level of disaggregation,
- increase systematically with temporal and cross-sectional aggregation, a finding consistent with that in the aggregation bias literature which assumes a Koyck model of the carryover effects of advertising (e.g., Clarke 1976; Russell 1988; Weinberg and Weiss 1982), and
- are not significantly different from zero when the data are at the most disaggregate level (i.e., panelists' brand choices), despite the fact that brands in the category have advertised heavily with wide variation over time and across brands.

These conclusions are based not only on the results reported here, but also on numerous other analyses that we do not report because of space constraints. In addition, the results are not affected by our data preparation procedures for two reasons. First, our results hold across different models, measures, and aggregation levels for which the same data processing was carried out. Second, we had previously processed the data differently, aggregating all variables to the brand level instead of the brand, size, and form levels as described here, and the pattern of results was very similar.

The generalizability of the results must await replications, but some of the results bear similarities to patterns in the literature. Our most important obser-

vation is that the estimated effects of advertising are significantly different from zero only when the data are aggregated over either time or individuals. In the literature, the level of aggregation is a key factor differentiating the two streams of research that led to the contrasting findings on advertising effects discussed previously. The older studies based on aggregate data estimated small but significant effects of advertising on sales (e.g., Assmus, Farley, and Lehmann 1983; Sethuraman and Tellis 1991). The newer studies based on disaggregate scanner data have estimated only weak or nonsignificant effects of advertising on brand choices (Kanetkar, Weinberg, and Weiss 1991; Tellis 1988a).

Which level of aggregation is more appropriate for measuring advertising's effects? An unambiguous answer must await formal analysis, but at least three arguments favor the more disaggregate level. First, sales response to advertising in categories such as the one we studied should occur as individual panelists are exposed to and process individual TV ads. Advertising dollars or gross rating points are at best proxies for the individual-level processing that must take place. Thus, the more the data represent this level of disaggregation, the more likely they will be to capture any relationship that might be present.

Second, for the preceding reason, econometricians generally believe that given different levels of aggregation, the more disaggregate data provide more consistent and less biased estimates (e.g., Train 1990, p. 2, 37-44). Specifically, data aggregation compresses the variance of individual variables and often introduces spurious covariation among pairs of variables

Table 3
Effect of Temporal Aggregation on Regression of Number of Choices on Marketing Variable
at Individual-Panelist Level

Independent Variables	Regression Estimates					
	Daily		Weekly		Monthly	
	β	t	β	t	β	t
Intercept	.04	6.8	.04	6.6	.04	5.9
Share voice _t (-)	.008	.4	.05	2.1	.05	1.6
Share voice _{t-1}	-.003	.4	.02	1.7	.07	2.6
Loyalty	.87	69.1	.88	69.2	.86	66.3
Price (-)	.03	2.8	.03	2.8	.03	2.7
Coupon	.10	17.1	.10	16.9	.09	15.3
Feature	.13	5.8	.12	5.2	.12	5.1
Display	.27	21.4	.27	20.5	.23	17.3

due to external events. For example, disaggregate scanner data contain great variance in exposure levels across panelists and across time within a brands' advertising schedule. Suppose individual TV exposures do not affect individual brand choices. In that case, aggregate advertising levels may relate to aggregate sales because managers target heavy buyers with more GRPs, or because managers increase advertising activity during the peak season. In general, data aggregation inflates covariation, leading to problems of multicollinearity and spurious causality; disaggregate data afford more precise estimates.

Which level of disaggregation is the best? Again, an unambiguous answer would depend on formal analysis. However, the differential influence of aggregation on the coefficients of advertising and the other variables probably offers a clue. For example, in Tables 3 and 4, why do the coefficients of advertising vary with aggregation whereas those of the promotion variables do not? The answer lies in understanding the distribution of these independent variables. Retailers normally do not change the promotion status of brands in the store within a week. Indeed, for most of the alternatives in the choice set, the promotion variables have the same status within the month, as only some of the brands go on promotion. The values at the daily level are therefore the same as those for the week and similar to those at the monthly level, so aggregation over those time intervals may not affect the coefficients. In contrast, TV advertising exposures

vary substantially by programs and days of the week. Thus, each time the data are aggregated, the distribution of these independent variables is affected, which in turn affects the estimated coefficients (Russell 1988).

A similar logic holds for cross-sectional aggregation. Ad exposures vary by household. Aggregation influences the distribution of that variable and thus its estimated effect on the dependent variable. But the promotional variables and especially coupons, as defined here, are common across households as they measure the availability of coupons in the market. Thus, cross-sectional aggregation does not affect the distribution of coupons and their impact on the dependent variable.

The most appropriate data level is therefore likely to be the level of aggregation that least disturbs the natural distribution of the independent variables. At that level their effect on the dependent variable can be estimated without bias. When the focus is the effects of TV advertising exposures, the most appropriate data levels seem to be the individual-household level and the daily (or lower if relevant) level.

Whether or not the arguments for disaggregated data hold depends on further empirical and theoretical research. Such research is urgently needed because of the vast expenditures on advertising and the fact that current managerial recommendations tend to be based primarily on analyses of aggregate data (e.g., Abraham and Lodish 1987, 1990).

Table 4
A. Effect of Temporal and Cross-Sectional
Aggregation on Regression of Number of Choices
on Lagged Share-of-Voice*

Time Period	Regression Estimates			
	By Panelists		By Market	
	β	t	β	t
Daily	-.003	.5	.03	2.8
Weekly	.023	1.7	.09	2.7
Monthly	.069	2.7	.21	3.3

B. Effect of Temporal and Cross-Sectional
Aggregation on Regression of Number
of Choices on Coupon*

Time Period	Regression Estimates			
	By Panelists		By Market	
	β	t	β	t
Daily	.10	17.1	.09	17.9
Weekly	.10	19.9	.10	17.4
Monthly	.09	15.3	.11	10.9

*Other variables not reported.

If disaggregate data and the results from choice models are valid, an important issue is why the effects of advertising on brand choice are so weak despite heavy expenditures by firms. For laundry detergents, our estimate is that firms spent more than \$250 million in one year. A related question is why firms continue to invest in advertising despite persistent doubts about its profitability raised by studies such as ours and others carried out over the years (Aaker and Carmone 1982; Abraham and Lodish 1990; Eastlack and Rao 1989). Possible reasons include efforts to support a premium price, to motivate the sales staff, or to motivate retailers to stock and promote the brand. Procedural reasons may be that brand managers set budgets as a percentage of sales or to match competitive advertising or on the basis of analyses of aggregate data and not by sales performance.

Whether these explanations are valid remains a subject for further research. However, even if they are valid, it is not clear that these explanations justify the historical level of advertising expenditures (Aaker and Carmone 1982; Sethuraman and Tellis 1991). Alternatively, it is not clear that the goals suggested by the explanations cannot be achieved more directly and efficiently by means other than advertising.

Another explanation sometimes given for the role of advertising is that it serves to enhance brand loyalty rather than to motivate current brand choice. Indeed, that is one of the most common hypotheses put forth in the industry. The evidence often provided for this hypothesis is that large-share brands have high loyalty and high advertising (e.g., Coke). However, such evidence is flawed by an incorrect test or interpretation of causality. For example, one could also argue that firms, successful because of superior quality, may win loyal consumers and a large share of the market. Such firms may advertise heavily to promote their quality or because they have the resources to do so. In such a case, quality leads to brand loyalty, market share, and heavy advertising, not the other way around. Other (more plausible) causes of brand loyalty are consumer inertia and market pioneering.

On the basis of our results, we question advertising's influence on brand loyalty because it appears to have so little effect on current brand choice. After all, if advertising does not influence consumers' selections of a brand at the time when its message is fresh, how will it affect a strong, enduring preference for a brand at a later time? Such an enduring preference is what we think marketers normally mean by the term "brand loyalty."

Some critics may argue that advertising works primarily by attracting new consumers to the market or the brand. That hypothesis is plausible and consistent with the view that advertising serves as a source of information for consumers. However, we question whether the proportion of new consumers in a mature category, such as the one we studied, is adequate to justify the heavy expenditures on advertising. We cannot draw definite conclusions on this issue, however, because the design of our sample did not include new consumers and the data did not include new products.

Conclusion

The urge to generalize from our results is strong. The data are detailed and the product class is well established, supported by heavy advertising for sev-

eral of its brands, and perhaps typical of many classes of packaged goods offered for sale in supermarkets. In addition, our results are typical of those reported in the literature for both choice and response models. Our study seems to us to be robust. Therefore, on the basis of the preceding discussion and the results from the literature, we are led to the working hypothesis that advertising plays a nondetectable role in the choice of brands by households in discrete time. The statistically significant effects of advertising found in studies using response models may be spurious results reached by aggregation of the data over time and households. Additional work is needed to either confirm or contradict the hypothesis.

References

- Aaker, David A. and James M. Carmone (1982), "Are You Over Advertising?" *Journal of Advertising Research*, 22 (4), 57-70.
- Abraham, Majid M. and Leonard M. Lodish (1987), "PROMOTER: An Automated Promotion Evaluation System," *Marketing Science*, 6 (2), 101-123.
- and — (1990), "Getting the Most Out of Advertising and Promotion," *Harvard Business Review*, (May-June), 50-58.
- Assael, Harry and David F. Poltrack (1992), "Using Single-source Data to Select TV Programs Based on Purchasing Behavior," *Journal of Advertising Research*, 31 (4), 9-17.
- Assmus, G., John U. Farley, and Donald R. Lehmann (1984), "How Advertising Affects Sales: Meta-Analysis of Econometric Results," *Journal of Marketing Research*, 21 (February), 65-74.
- Bass, Frank M. and Robert P. Leone (1983), "Temporal Aggregation, the Data Interval Bias, and Empirical Estimation of Bimonthly Relations From Annual Data," *Management Science*, 29 (January) 1-11.
- and — (1986), "Estimating Micro Relationships From Macro Data: A Comparative Study of Two Approximations of the Brand Loyal Model Under Temporal Aggregation," *Journal of Marketing Research*, 23 (August) 291-297.
- Ben-Akiva, Moshe and Steven R. Lerman (1989), *Discrete Choice Analysis*, Cambridge, MA: The MIT Press.
- Clarke, Darral G. (1976), "Econometric Measurement of the Duration of Advertising Effect on Sales," *Journal of Marketing Research*, 13 (November), 345-357.
- Cooper, Lee G. and Masao Nakanishi (1988), *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*, Boston: Kluwer Academic Publishers.
- Curry, David J. (1993), *The New Marketing Research Systems*, New York: John Wiley & Sons, Inc.
- Eastlack, Joseph O., Jr., and Ambar G. Rao (1989), "Advertising Experiments at the Campbell Soup Company," *Marketing Science*, 8 (Winter), 57-71.
- Griliches, Z. (1967), "Distributed Lags: A Survey," *Econometrica*, 35 (January), 16-49.
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-238.
- Gupta, Sunil (1988), "Impact of Sales Promotion on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342-355.
- Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, 2nd ed., New York: John Wiley & Sons, Inc.
- Kanetkar, Vinay, Charles B. Weinberg, and Doyle L. Weiss (1992), "Price Sensitivity and Television Advertising Exposures: Some Empirical Findings," *Marketing Science*, 11, (4), 359-371, 1992.
- Lambin, Jean J. (1976), *Advertising, Competition and Market Conduct in Oligopoly Over Time*, Amsterdam: North-Holland Publishing Company.
- Lattin, James M. and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299-310.
- Leone, Robert P. and Randall L. Schultz (1980), "A Study of Marketing Generalizations," *Journal of Marketing*, 44 (1), 10-18.
- McFadden, Donald (1974), "Conditional Logit Analyses of Qualitative Choice Behavior," in *Frontiers in Econometrics*, P. Zarembka, ed., New York: Academic Press, 105-142.
- Palda, Kristian S. (1964), *The Measurement of Cumulative Advertising Effects*, Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Pedrick, James H. and Fred S. Zufryden (1991), "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single-Source Data," *Marketing Science*, 10 (1), 111-130.
- Russell, Gary J. (1988), "Recovering Measures of Advertising Carryover From Aggregate Data: The Role of the Firm's Decision Behavior," *Marketing Science*, 7 (3), 252-270.
- Sethuraman, Raj and Gerard J. Tellis (1991), "An Analysis of the Tradeoff Between Advertising and Price Discounting," *Journal of Marketing Research*, 28 (May), 160-174.
- Simon, Julian L. and Johan Arndt (1980), "The Shape of the Advertising Response Function," *Journal of Advertising Research*, 20 (4), 11-28.
- Sims, C. (1971), "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 39 (May), 543-563.
- (1972), "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of American Statistical Association*, 67 (March), 169-175.
- Tellis, Gerard J. (1988a), "Advertising Exposure, Loyalty and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25 (May), 134-144.
- Tellis, Gerard J. (1988b), "The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales," *Journal of Marketing Research*, 25 (November), 331-341.
- Train, Kenneth (1990), *Qualitative Choice Analysis*, Cambridge, MA: The MIT Press.
- Weinberg, Charles B. and Doyle L. Weiss (1982), "On the Econometric Measurement of the Duration of Advertising Effect on Sales," *Journal of Marketing Research*, 19 (November), 585-591.
- Weiss, Doyle L., Charles B. Weinberg, and Pierre M. Windal (1983), "The Effects of Serial Correlation and Data Aggregation on Advertising Measurement," *Journal of Marketing Research*, 20 (August), 268-279.
- Windal, Pierre M. and Doyle L. Weiss (1983), "An Iterative GLS Procedure for Estimating the Parameters of Models with Autocorrelated Errors Using Data Aggregated Over Time," *Journal of Business*, 53 (4), 415-24.