The Effects of Television Advertising on Local Telephone Usage: Exploratory Data Analysis and Response Modeling

Author(s): Martin A. Koschat and Darius J. Sabavala


Published by: INFORMS

Stable URL: http://www.jstor.org/stable/183849

Accessed: 12/11/2013 11:40

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.
THE EFFECTS OF TELEVISION ADVERTISING ON LOCAL TELEPHONE USAGE: EXPLORATORY DATA ANALYSIS AND RESPONSE MODELING

MARTIN A. KOSCHAT and DARIUS J. SABAVALA

Yale University
Polytechnic University

In increasingly competitive and deregulated service environments, telecommunications service providers are interested in using advertising to stimulate usage. Measuring response in telephone usage to advertising presents difficulties because the effects are likely to be small, and the usage exhibits high variability. The substantive question of whether advertising has any effect on telephone usage is addressed through a systematic analytical process that combines exploratory data analysis with formal modeling. We find that telephone usage does respond to advertising, that this response can be quantified, and that households with heavy usage of telephone service respond proportionally more than light usage households.

(Advertising and Media Research; Buyer Behavior; Estimation and Other Statistical Techniques; Regression and Other Statistical Techniques; Telecommunications.)

1. Introduction

Does advertising designed to stimulate consumption and purchasing work? Accurate, efficient measurement of customer transactions, in natural settings and partially controlled experiments, especially for frequently purchased consumer packaged goods, has significantly raised expectations about our ability to discern customer response to marketing efforts. Although more data of better quality are available and advances in the analysis and modeling of such data continue to be made, researchers continue to grapple with this basic question.

In this paper we address the challenge of measuring advertising response in a services market. The primary objective of the local telephone company, which has conducted the experiment we analyze, is to determine the potential for stimulating demand for local telephone usage through advertising in the presence of the considerable advertising activity of long-distance companies. Several aspects of our problem setting appear ideal for the purpose of assessing advertising effects, including the lack of competition, the absence of price effects, and the potential for immediate response.

Measuring response in telephone usage to advertising presents a great challenge because the effects are likely to be small, and the usage (at individual and aggregate levels) exhibits

1 Data and results have been scaled and some numerical values suppressed to protect their proprietary nature while preserving the necessary information to illustrate our analytical results and support our substantive conclusions.
high variability. The published research in advertising, including reviews by Little (1979) and Aaker and Carman (1982), and the experimental studies of Ackoff and Emshoff (1975), Carroll et al. (1985), Haley (1978), acknowledges the difficulties encountered in measuring the effects of advertising. McGuire (1977) and Winer (1980) address the analysis of advertising experiments.

More specifically, the problems begin with the constraints posed by experimental design and center quite heavily on the extremely high customer heterogeneity that is typical of consumption behavior. Kuritsky et al. (1982) encounter problems akin to ours in a very similar application.

To deal with these problems systematically we address the following tasks:

i. Develop parsimonious descriptions of customers’ usage profiles.

ii. Compare the test and control customer panels on their usage profiles.

iii. Relate usage levels to customer characteristics such as demographics.

iv. Select a measure of response that is plausible on behavioral grounds, has tractable statistical properties, and permits the exploration of how usage response depends upon prior usage levels.

v. Specify a model for this response measure that incorporates the effects of prior usage levels as well as demographics.

vi. Refine this model to yield quantitative information on the magnitude of the response and to identify customers who are more responsive.

Challenged by the constraints of a typical (not quite textbook) experiment and the complex patterns of telephone usage, our analytical process combines exploratory data analysis and formal modeling. If the advertising-usage response mechanism and the stochastic properties of usage are well understood, then formal modeling may proceed directly. In the absence of generally accepted advertising theories, we propose that exploratory analysis of the data is an essential precursor and complement to formulating, calibrating, and refining an advertising response model.

Graphical analyses are used, not only to present results for visual confirmation, but, with other exploratory and model-based techniques, also to seek structure in the data and provide diagnostic tools. Box (1988), in the context of quality assurance, stresses the importance of exploratory analysis in the early stages of scientific investigation:

A first analysis of experimental results should, I believe, invariably be conducted using flexible data analytical techniques—looking at graphs and simple statistics—that so far as possible allow the data to 'speak for themselves'. The unexpected phenomena that such an approach often uncover can be of the greatest importance in shaping and sometimes redirecting the course of an ongoing investigation.

These ideas by themselves are not new. (See Tukey (1977), Mosteller and Tukey (1977), and Chambers et al. (1983).) In fact, most of the techniques used here have been extensively tried and tested theoretically and empirically. Our contribution lies in applying these ideas for measuring advertising effects on local telephone usage, and in conjunction demonstrating their value in guiding formal modeling. To be truly useful as a supplement to formal modeling such techniques must be readily available to academics and practitioners. The techniques used in this paper are essentially available in major statistical packages on various computing platforms. Our analysis relied on S (see Becker et al. (1988) and Statistical Sciences Inc. (1990)).

A glimpse at our results ratifies our decision to integrate judgmental and formal methods—local telephone usage does respond to advertising, and this response can be quantified. Furthermore, the individual advertising response depends upon the individual average level of telephone usage. Therefore, in general, a greater response can be attributed to heavy users, and a small proportion of telephone users contribute much of the response.

In the next section we describe the telecommunications problem setting, the experimental design, and the data available. We are driven by the importance of the study
objectives to the local telephone company, and our analysis seeks to confront the technical issues present. In the third section we describe the results of the exploratory analysis, their substantive meaning, and the implications for model specification. We close this section with a specification of an individual response model. The following section applies and refines the model and generates the final results to our basic question. The paper concludes with the implications of our study for telecommunications industry managers, especially in local telephone companies, and a discussion of the nature of the behavioral response found in comparison with other findings in the literature.

2. The Telecommunications Setting

2.1. Telecommunications Advertising and Experimental Design

In 1984, the Bell Telephone system in the United States was reorganized as a result of an antitrust settlement decree into seven regional companies and AT&T. Each regional company, comprised of operating telephone companies, provides local access to the telephone network and calling services within specific geographical areas where they operate as mostly exclusive franchised carriers. AT&T and other long distance carriers like MCI and U.S. Sprint compete for service traffic between these geographical areas.

Before 1984, the Bell operating telephone companies shared in the revenues from the profitable long-distance market. After 1984 the revenues for these companies from long-distance have greatly diminished, resulting in a desire to develop their business within their regions. Advertising is an increasingly important variable in the marketing program to stimulate demand. Since long-distance companies have operated in a competitive environment they have engaged in considerable advertising for usage stimulation. In most cases, the advertising messages have sought to stimulate customers’ usage of the telephone in general, rather than calls of one type or another. If customers respond to a long-distance telephone company’s advertising by making incremental long-distance (more expensive) calls, then they are likely also to respond by making incremental local (less expensive) calls, subject of course to the customers’ community of interest. Therefore the management of a local telephone company must determine whether their own advertising has an incremental effect on local usage beyond that caused by the long-distance advertisers. The experiment that we analyze was designed and conducted by a local telephone company that sought to determine whether advertising could stimulate local telephone calling by their residence customers, to identify which customers were most responsive, and to measure the degree of stimulation.

Compared to a long distance carrier, the local telephone company is at a disadvantage in addressing this question since its experience with advertising is quite recent, and since the opportunity to conduct highly controlled experiments within their region is more limited. It is worthwhile to compare this study with one of the few published studies that are directly relevant. Kuritsky et al. (1982) analyzed an experiment designed to study the effects of advertising and advertising copy on long-distance telephone calling. Although their market provided larger samples and a test of longer duration, the modeling challenges they encountered parallel ours. We refer to these as they occur in the rest of the paper.

The advertising control was achieved using split-cable television. Split-cable television is an unobtrusive means of controlling the television commercial exposure of a panel of households in the same city. Households in the same neighborhood may be assigned to either of two cables, enabling the advertiser to send different campaigns (or none) to each of the groups. The assignment is intended to produce groups that are essentially similar in aggregate, despite differences among households. The comparability of groups used in the experiment will be assessed below. A special sample of 800 residential customers with single-line telephone service were randomly selected from each of two cable groups.
The experiment consisted of placing a series of advertisements on the cable of one of the groups, the test group, and placing public service announcements (no advertising) on the cable of the other group, the control group. The campaign used two messages, one stressing the time-saving benefits of calling, while the other stressed the emotional rewards of calling. The television commercials for the first message showed several situations in which the protagonist would have saved traveling, found the right shop, obtained an appointment, and avoided embarrassment by telephoning in advance. The copy for the second message portrayed situations in which the protagonist could call to make the recipient of the call happier or less anxious; for example, calling their mother who is alone, calling their parents when they are late, calling the children when they work late, calling after a lovers' quarrel. Since this was not intended to be a copy test, these different messages were mixed together in the sequence of test commercials and the entire test group received the same mix. Following the previous positioning for this staple service, the target segment was broadly defined to be adults in all households.

The media schedule attempted to achieve a stable GRP level of 300 per week corresponding to an average exposure level of 3–5 per week over the trial period of 38 weeks. During the pretrial period of 13 weeks, neither group received any local usage stimulation advertising. Weekly expenditures and the number of inserts were recorded, but the actual GRPs were not. In addition, the technology being used did not permit the gathering of household viewing data.

2.2. Customer Level Data

The customer level data consisted of customer demographics and detailed records of local telephone usage.

Demographic data for the sample were collected using a survey and included number of telephone sets, reasons for calling (percentage of calls for social/personal/community-service/business), perceived value of the telephone, type of home, years in home, household size and composition by age group, age of head of household, employment and education of male and female heads, and income. Of the 14 original variables, some were in categorical form and are converted to suitable dummy variables. The result is a set of 30 demographic variables for analysis, several of which are indicators.

Telephone usage data for each member household in the test and control groups covering the pretrial and trial periods were obtained from the company's internal records. The extracted data included the weekly frequency and aggregate duration of calls. While the experiment started with 1,600 households, incomplete data resulted from nonparticipation in the experiment, survey non-response (missing demographic data), termination of cable service (exclusion from the advertising control), or termination of telephone service (probably due to relocation, exclusion from the experimental sample). These households were dropped. Finally, elimination of a few households which made no calls resulted in a sample of 360 households in the test group, and 375 households in the control group.

To summarize, our data at the household level consist of demographics and telephone usage data for 13 pretrial weeks and 38 trial weeks for each of 360 test households and 375 control households.

2.3. The Measurement and Modeling Challenges

Experience in modeling purchasing in many contexts has pointed to the considerable influence of customer heterogeneity. For our problem, Figure 1 shows that the differences between average trial period usage and average pretrial period usage vary substantially from customer to customer. This high level of variation makes it difficult to discern an advertising effect. In addition, it is clear that both distributions are non-normal. While visual inspection appears to suffice in this case, we confirmed the non-normality using
a Q-Q plot, a graph of the quantiles of the empirical distribution against corresponding quantiles of a comparison distribution, the theoretical normal in this case (see Chambers et al. 1983). Applying an ANOVA model, with its standard distributional assumptions of normality and homoscedasticity, will yield statistically inefficient estimates and may lead to erroneous conclusions.

Direct approaches might be informative in situations where the customers are relatively homogeneous, the response is overwhelming, and dynamic effects (trend and seasonality) are minimal. Unfortunately, such conditions do not prevail here. These basic threats to the validity of conventional statistical approaches are well recognized, although the solutions are more elusive; see Cook and Campbell (1979). Anticipating these problems, our approach for measuring the advertising effects rests on developing a sound understanding of individual level usage patterns. The results are used to guide us in developing an appropriate response measure and a model specification at the individual level.

3. Development of a Model of Individual Response

Modeling individual response has received considerable attention in recent years with the availability of customer-level data. In our analysis of the experiment, we seek to determine the magnitude of the change in usage due to advertising and an explanation that can be supported by experience as well as by a judicious examination of the data. As with any modeling task, specification requires the selection of appropriate variables, measures of those variables, and an appropriate functional form. The analysis then entails estimation and refinement of the model specification, tasks made easier and potentially more fruitful by suitable diagnostics. Throughout this process we used exploratory data analysis techniques, most of which are well established but some of which have been developed more recently.

3.1. Characterizing Individual Usage

Usage is complex and multidimensional. Given our data, each customer record can be viewed as an observation from a 51 variable (13 plus 38 weeks) distribution. Using principal components analysis, we seek a low-dimensional parsimonious characterization of usage that captures the essential elements of the customers’ usage profiles. By doing so we reduce the complexity of measuring and analyzing response due to advertising.
Figure 2(A) shows the distribution of weekly calling frequencies, reflecting the extreme heterogeneity of customers. The number of calls per week range from 0 to 256 with a mean of 35 and a median of 24. Such extreme skewness presents problems in scaling (across individuals), difficulty in detecting outliers, and difficulty in using standard statistical machinery. To address these problems we transform the data so as to permit the application of principal components analysis and related testing procedures. Following the work of Pavarini (1979), a power transformation of 0.25 leads to a marginal distribution which is close to normal, as seen in Figure 2(B), and confirmed by a Q-Q plot. Visual inspection using advanced dynamic graphical visualization tools (see Koschat and Swayne 1993) suggests that the joint distribution is also close to normal. As a result, the power transformation applied to averages across weeks for each individual will also yield normality.

As a practical matter, we performed separate principal components analyses for the 13 variables of the pretrial period and the 38 variables of the trial period. Test and control groups were analyzed jointly for the pretrial period but separately for the trial period, resulting in three distinct analyses. Since the findings for all three analyses are almost identical, only the ones for the pretrial period are described in detail. Based on the sample covariance matrix of the transformed data, the table below shows the percent of total variation explained by the first \( k \) principal components and the Akaike Information Criterion (AIC = \(-2 \) maximum log-likelihood + 2 Number of parameters) for a model that assumes that the 14-\( k \) smallest eigenvalues of the covariance matrix are equal (see Akaike 1987).

<table>
<thead>
<tr>
<th>( k )</th>
<th>Akaike Information Criterion</th>
<th>Cumulative % Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>115699</td>
<td>75.5</td>
</tr>
<tr>
<td>2</td>
<td>5689</td>
<td>80.5</td>
</tr>
<tr>
<td>3</td>
<td>5247</td>
<td>84.0</td>
</tr>
<tr>
<td>4</td>
<td>4742</td>
<td>96.9</td>
</tr>
<tr>
<td>5</td>
<td>4741</td>
<td>97.0</td>
</tr>
<tr>
<td>6</td>
<td>4744</td>
<td>99.0</td>
</tr>
<tr>
<td>7</td>
<td>4746</td>
<td>100.0</td>
</tr>
</tbody>
</table>
The first principal component clearly dominates the remaining 12 components by explaining 75.5% of the total variation. Inspection provides a straightforward interpretation of this principal component; it corresponds to each household's average usage across the 13 weeks of the pretrial period. Despite this dominance of the first principal component, the AIC suggests as many as 10 distinct eigenvalues. Most of the subsequent usage analysis focuses on the first principal component, that is, average usage. We replicated our entire analysis using higher order components, but did not discover any interesting or relevant relations.

The average number of calls is an excellent, interpretable characteristic of usage that can be used to differentiate customers. With this characterization, given the presumably random allocation of households across the dual cables, we have a basis for assessing the degree to which the distributions for test and control are matched. If they are closely matched in the pretrial period then we can compare the distributions in the trial period to measure the effect of the advertising. An empirical Q-Q plot comparing the test and control distributions in the pretrial period is shown in Figure 3(A). The two groups are fairly well matched except for the upper tail; this discrepancy is less pronounced for the transformed pretrial average.

If the heavier users, who account for a disproportionate share of the usage, also account for a disproportionate share of the response, then a measure of the experimental advertising effect obtained from a simple comparison of test and control is likely to be biased and possibly inefficient. Kuritsky et al. (1982) encountered the same problem and chose to measure response (average trial usage–average pretrial usage) separately for segments of customers. Our approach is similar in spirit although we continue to work with individual customers and focus explicitly on uncovering how the response relates to usage characteristics.

3.2. Relating Usage to Demographics

To correctly interpret a model relating response to usage characteristics, we need to ensure that a measured differential advertising response based on usage level is not due to some other confounding variable. Accordingly we need to understand whether usage depends upon other customer characteristics, such as demographics, a topic of interest in its own right. A notable study for comparison is that of Brandon (1981) who analyzes the impact of demographics on local telephone usage in the Chicago market.
A regression of the power-transformed pretrial average usage on the 30 demographic variables resulted in an $R^2$ of 0.05, suggesting little or no explanatory power. However, if we judiciously aggregate and select demographic categories based on the findings in Brandon (1981), then we find that telephone usage levels increase with the presence of children under 10 years of age, household size, and the head of household being retired. Brandon found a positive effect of high income but we find no such effect, possibly due to the absence of price effects in the current study. The $R^2$ in Brandon’s study was higher, 0.20, perhaps as a result of some differences: the data in Brandon’s study were from Chicago proper (possibly a more heterogeneous population), local service was charged based on usage (permitting income to have an influence), and race had the greatest predictive power (this variable was not available in our study). Overall though, the results bear a reasonable resemblance to those of Brandon. What is more notable and common is that the link between demographics and usage is extremely weak.

3.3. Choice of Response Measure

Our choice of the appropriate response measure, the dependent variable in our model, is made on conceptual and statistical grounds. The measure should reflect a plausible behavioral response to advertising and should be interpretable. Also, the distributional properties of the measure, in particular skewness and multi-modality (which make modeling difficult), should be tractable.

Since the test and control groups are not closely matched on usage, the response measure is based on trial usage compared with pretrial usage. Given that customer heterogeneity with respect to usage is a major source of variation, we choose to use the ratio (trial/pretrial). Such a measure implicitly includes seasonality, trend, and other effects. However, these will be common to test and control households.

The ratio is the preferred response measure for several reasons. It is interpretable (proportional response) and accounts for the behavioral response to advertising (heavy users can respond more than light users in absolute terms). In addition, the ratio has nicer distributional properties. Characterizing each customer by their pretrial and trial averages, plots of the ratio (trial/pretrial) against the power-transformed pretrial mean are shown in Figure 4, separately for test and control groups. We also considered the difference (trial – pretrial) and found that the difference, conditional on the power-transformed pretrial average, increases considerably in variance with the pretrial average. As seen in Figure 4, the ratio has a more stable variance.

We are interested in the relation between the response and the pretrial usage as seen in Figure 4. The functional form of the conditional expectation of the ratio given the pretrial average will be useful in specifying the functional form of the response model, especially since demographic variables are not closely associated with usage. Using LOWESS (Cleveland 1979), a non-parametric regression tool, we estimate a flexible form of the conditional expectation. The result, seen as the solid lines in Figure 4, indicates nonlinearity. We approximate this nonlinear relation by a piecewise linear function with a kink at around 19 calls per week per household (corresponding to a power-transformed weekly average of 2.15). Sensitivity analysis of our critical results showed considerable stability with regard to the choice of the kink.

Two points are worth noting. First, the slopes above the kink for both the test and the control group are clearly negative, which could possibly be an indication of regression to the mean. In the presence of uncertainty in the data, the trial mean conditional on a specific pretrial value will be closer to the mean level than the pretrial value is to the pretrial mean (hence the concept of “regression to the mean”). The greater the distance of the pretrial value from the pretrial mean (as would be the case for very heavy users), the larger is this “regression” effect; see Nesselroade et al. (1980). Second, while there is no discernible difference between the test and control groups in the slopes below the
kink, there appear to be differences above the kink. Essentially, the decrease from pretrial to trial (the level of the ratio is less than one) is greater for larger users, and the slope is less negative for the test group. Whether this difference in slopes can be attributed more to the experimental effect rather than an effect related to the regression to the mean is an issue we will explore in the context of a specific model.

3.4. Specification of A Model of Individual Response

The results we have obtained above are interesting in and of themselves, and they are also important because they guide the specification of the model in a direct way. In brief, we model the ratio response as a function of a piecewise linear component of usage, demographic variables, and demographic/advertising interaction variables.

Denote by $r$ the ratio of a customer’s weekly average usage for the trial period and the one for the pretrial period, by $m$ the power-transformed pretrial weekly average, and by $m_0$ the kink in the piecewise linear function which we chose to be 2.15. The variables $I_t$, $I_l$ and $I_h$ are indicator variables that take on values of 0 and 1. Variable $I_t$ equals 1 for the test households, $I_l$ equals 1 for $m \leq m_0$, and $I_h$ equals 1 for $m > m_0$. The $\alpha$ and $\beta$ parameters denote intercepts and slopes. Subscripts of $t$ and $c$ refer to the test and control households, while superscripts of $l$ and $h$ refer to low usage and high usage households (with $m$ below and above $m_0$, respectively). Also define $d$ as a row vector of the 30 demographic variables for a given household. Let $\gamma$ be a column vector representing coefficients of $d$. Then, a formal statement of the model for a given household is:

$$E(r|m, d) = \alpha_t + \beta_t^l \times I_l \times (m - m_0) + \beta_t^h \times I_h \times (m - m_0) + (\alpha_t - \alpha_c) \times I_t + (\beta_t^l - \beta_t^c) \times I_l \times I_t \times (m - m_0) + (\beta_t^h - \beta_t^c) \times I_h \times I_t \times (m - m_0) + d \times \gamma_c + I_t \times d \times (\gamma_t - \gamma_c).$$

This model has six parameters for the piecewise linear functions for test and control households, and sixty parameters for the demographics and demographics/trial inter-
actions. The effect of advertising is reflected in the differences between the test and control line segments, and for this reason, the variables are specified so that no difference (between test and control) is reflected in zero coefficients of the corresponding variables.

4. Diagnostics, Model Refinement & Model Selection

Model selection and inference critically depend upon whether the distributional assumptions are tenable. As a result, it is appropriate to refine the model based on diagnostics from an initial estimation. Two key refinements are deemed to be necessary. The first is that a moderate skew persists in the ratio (seen in the residuals to the initial estimate of the full model). The obvious solution of using the ratio of power-transformed averages corrects the skewness. The second is that heteroscedasticity with respect to the usage level (pretrial average) continues to be a problem. Direct estimation of the standard error of the residual, conditional upon the usage level, as described in Appendix A, shows that the variance is greatest at the extreme usage levels, with a minimum at a (transformed) usage level of 2.8. We make a rough correction to the heteroscedasticity via a weight based on the inverse of the estimated variance.

The objective of model selection is to reduce the full model of 66 variables to a more parsimonious form. We identified several candidate models using a variety of approaches, including forward and backward stepwise selection, minimizing the previously introduced Akaike Information Criterion (AIC), and using the closely related $C_p$ (Mallows 1973). All of these candidate models included the intercept, the slope below the kink for the control group, the slope above the kink for the control group, and the slope differential above the kink for the test group. The other variables in each of the candidate models are three to five demographic and demographic/trial variables. None of the demographic variables withstood further scrutiny. For example, imposing the sensible restriction that a demographic/trial interaction variable be only included if the corresponding demographic main effect be included also, the model with the minimum AIC has three demographic indicators. By themselves, none of these variables is statistically significant (coefficient different from 0 tested via the coefficient’s t-statistic) at the .95 level, nor are they significant as a group (incremental variance explained as a group from an appropriate F-test at the .95 level).

With the demographic variables out of the way (neither exclusion nor inclusion of demographic variables has a significant effect), we focus on finalizing the specification of the advertising-usage function. Starting with a common piecewise linear model for test and control we find no incremental explanatory power by including the trial effect, that is, the main effect of advertising ($p$-value of appropriate $F$-test = 0.14). Adding the slope differential above the kink to the base model or to the base model with the trial effect adds significant explanatory power ($p$-values of 0.01 and 0.03). Finally, adding the left-hand slope differential to either of these intermediate models is not worthwhile ($p$-values of 0.22 and 0.23). In all cases the slope differential above the kink does not change much in value. As a result we conclude that:

$$E(r^{1/4}|m, d) = \alpha_c + \beta_i^c \times I_t \times (m - m_0)$$

$$+ \beta_h^c \times I_h \times (m - m_0)$$

$$+ (\beta_i^h - \beta_i^c) \times I_t \times I_h \times (m - m_0).$$

The results of weighted least squares estimation with the weights of Appendix A are found as (the results for the intercept are not reported to protect the proprietary nature of the data):
The difference between test and control households, presumably due to advertising, is
carried to the slope for the heavier users. The advantage of using weighted least squares
is that a confidence interval for the parameter of interest, $\beta_i^h - \beta_i^c$, is easily obtained.
However, the estimated standard error of the parameter is conditional on the weights
that are used. Incorporating the contribution to uncertainty due to the weighting cal-
culation would be extremely difficult, if not impossible. Accordingly we seek to examine
the validity of the weighted least squares estimate of the standard error of the coefficient
by performing a bootstrap to compute the same quantity. The bootstrap method and the
results are described in Appendix B. The weighted least squares estimates are seen to be
sufficiently accurate, and we conclude that the parameter $\beta_i^h - \beta_i^c$ is indeed positive.

Does this mean that the advertising response increases with an increase in the pretrial
usage level? Consider the alternatives of an uniform advertising response. The conditional
expectation of the ratio for the control group is

$$r_c(m) = E(M_{\text{trial}}/M_{\text{pretrial}} | M_{\text{pretrial}} = m).$$

Under an additively uniform advertising response the conditional expectation of the
ratio for the test group would be

$$r_t(m) = E((M_{\text{trial}} + a)/M_{\text{pretrial}} | M_{\text{pretrial}} = m) = r_c(m) + a/m.$$  

Under a multiplicatively uniform advertising response the conditional expectation of
the ratio for the test group would be

$$r_t(m) = E(A \times M_{\text{trial}}/M_{\text{pretrial}} | M_{\text{pretrial}} = m) = A \times r_c(m).$$

With $a > 0$, $A > 1$ and $r_c(m)$ decreasing in $m$, it is $r_t(m)$ rather than $r_c(m)$ that has
the steeper negative slope both under the additive and the multiplicative model of uniform
response. Since we observe a more positive slope for the test group we may indeed conclude
that an increase in advertising response is associated with an increase in pretrial usage.

5. Discussion

5.1. Implications for Telecommunications Advertising and Experimentation

In the presence of heightened competition among long-distance companies, local tele-
phone companies must determine whether the allocation of their resources to advertising
can have an incremental effect on local usage. From the perspective of the telephone
company involved in this study, the most important finding is that advertising is seen to
have a significant impact on telephone usage.
The study focused on local telephone usage, offered for a fixed monthly fee that covered an unlimited number of calls. Since there is no incremental charge for a local call, the number of calls for any customer is constrained only by the value of the call and by the opportunity cost of the customer's time. The decision to focus on local calling was motivated by the greater volume of local calls relative to toll calls. From an analytical viewpoint, the prevalence of a strong Poisson component in telephone usage supports the use of the higher volume service in the experiment since the coefficient of variation decreases with an increase in the mean. On the other hand, customers who make additional calls do not incur any incremental charges, so that these calls do not increase revenues for the telephone company. The rationale rested on the limits imposed by where the study might take place, along with a managerial imperative to examine the behavioral impact of advertising free of price effects. If there are no effects at zero prices, there are unlikely to be effects at non-zero prices.

Calling frequency, of course, is not the only measure of demand. Telephone usage, like other transaction-oriented services, may be measured on several other dimensions including the durations of the calls, the distance called, the time of use, and the type of services and features being used (for example, operator assistance). In our example, the customer usage data included the duration of the calls. A parallel analysis showed that there were no discernible effects of advertising on the average duration of calls. The implication for managers is that promotional efforts may induce a behavioral response that varies across the usage dimensions, with revenue effects that depend upon the price structure. As discussed below, the nature of the response may be influenced by the advertising message.

This experiment demonstrates the benefits of market experimentation while also highlighting its difficulties. These difficulties fall broadly into three categories: (1) choice of the experimental test site, (2) decisions which customer responses to record, and (3) creation and maintenance of a database of customer responses. With these difficulties, the resulting experimental design reflects several compromises, which, in turn, impose limitations on the analysis.

The choice of the market for the experiment is severely constrained by the availability of advertising test locations with split cable technology within the operating region of the telephone company. In addition, there are constraints on the test and control locations where local telephone usage can be monitored. In our study these considerations left only one test market. This then begs the issue of how transferable the results might be from the test market to the entire market. An economic evaluation of advertising in the entire operating region of the telephone company would further need to take account of mitigation in message volume response due to different prices and inclusion of revenues generated by a response in non-local calls.

In principle, a telephone company's internal records provide extensive customer-level data, with the potential to obtain large and well-designed samples. This is true not only in telecommunications but in many other industries, for example, other utilities and financial services. Such a data collection effort is beset with three kinds of problems. First, the databases are usually not designed for marketing analysis—if the data are not essential for operations or billing, they are not recorded. In fact, in this experiment a special effort was needed to measure usage, since due to the flat-rate local pricing scheme, local calls were not metered prior to the experiment. Second, even if the data are recorded, historical data may not be saved or be aggregated losing detail. Third, even if the data

2 Historically, local calling has been priced in a bundle with access to the network, and therefore, has not been charged for on a usage-sensitive basis. Several operating telephone companies have introduced, and continue to introduce, pricing structures that charge for local calls, subject to regulatory approval from the public utilities commission.
are saved, the extraction and handling of a very large database from a much larger one is time-consuming and costly, with the potential for further errors. The benefits derived from experimentation of the kind described here clearly suggests that the telephone company add the flexibility to its billing system required to facilitate data extraction. As a matter of fact, a special effort is made to add such flexibility to the next generation of billing systems.

5.2. Nature of Advertising Response

A major focus of our study is to understand the nature of customer response to advertising. From a substantive point of view, the most important finding is that the degree of the response increases with the average usage level, and in fact, the response is due almost entirely to the heavier users.

It is instructive to think of possible explanations for the interaction between advertising response and usage level. There are several candidates and the potential for their relevance in this study. We conceive of two structural explanations.

i. Heavier users could have been targeted by the media plan and received more exposures. Our understanding of the media selection and planning for the experiment is that this explanation is extremely unlikely.

ii. Heavier users might have a lower opportunity cost of time, partly evidenced by their heavy usage in the first place. This, too, is less likely, since it would suggest that high income households, who we would expect to have a higher opportunity cost, made fewer calls or responded less to advertising; no such effects were found.

We consider four possible behavioral explanations.

iii. It may be that heavier users are more attentive to the advertising given their greater interest in the product. Such selective exposure has been studied by Silk and Geiger (1972). While selective exposure has not been definitively linked to purchase response, it is a concept that is widely acknowledged and consistent with our findings.

iv. Heavy users have more “opportunities to stimulate.” Such opportunities may be a consequence of the greater number of applications for which the product is used by heavy users and may also be due to the greater creativity evinced by heavy users in finding applications. This is clearly consistent with our findings.

v. Another possible behavioral explanation is that of “framing,” which is advanced by Deighton et al. (1993), whereby advertising interprets and enhances recent consumption experience. This enhanced experience leads to greater perceived value, which, in turn, leads to increased usage.

While this idea has intuitive appeal in general, in our specific case it would have to occur more frequently for heavy users. Referring back to the previous hypothesis regarding opportunities to stimulate, if heavy users as a group had a more extensive portfolio of applications and potential applications than the group of light users, then they would have proportionally more occasions in which the framing might occur. Viewed in this manner, framing would be consistent with our results.

vi. As suggested and supported by Raj (1982), it may be that loyalty influences response. Such an explanation would also be consistent with the findings in Guadagni and Little (1983) and Tellis (1988). To the extent that loyal purchasers are heavy users, promotional efforts including advertising may reinforce their purchasing behavior, and therefore this explanation would be consistent with our findings. In our setting, since there are no competitive brands, we would take loyalty to mean a tendency to use telephone calling instead of alternate forms of communications (for example, a personal visit or a letter).

A consideration that diminishes the loyalty explanation is that if loyalty effectively leads to repeat purchasing (as it should) then the effect of past purchases may dominate any effects of advertising. Deighton et al. suggest and find such an effect in the context
EFFECTS OF TELEVISION ADVERTISING ON LOCAL TELEPHONE USAGE

of brand usage for consumer packaged goods, which they term “usage dominance.” Therefore, heavy users of a brand would rely more on their experience (as opposed to advertising) to make a choice, and we would tend not to see the results we did.

Our findings viewed in context of the product and the consumption setting are potentially consistent with any or all of these behavioral explanations.

5.3. Nature of Copy Effects

While the advertising copy was not part of the experiment, it is recognized as an important conditioner of response. Some advertising may produce category level effects, while other advertising may work for specific products. For example, advertising the emotional benefits of telephone calling potentially would stimulate calls at all distances determined by the caller’s community of interest, while advertising that included prices for specific calls would potentially have an effect only on the customer’s demand for those calls.

In the context of copy effects, it is interesting to compare our findings with the findings by Kuritsky et al. (1982), who report the results of a test conducted by AT&T in the late 1970s comparing the differential benefits of its “Reach Out” and “Cost of Visit” campaigns on long distance calling. The creative theme of the “Reach Out” campaign with its emphasis on the emotional benefits of telephone calling is in many respects very similar to the creative theme used in the advertising of this study. The “Cost of Visit” campaign was designed to modify incorrect perceptions of the cost of long distance calling among light users and included a strong price message. Kuritsky et al. report a positive difference between the additional calls generated by the “Cost of Visit” campaign and the calls generated by the “Reach Out” campaign. They also report that this difference is larger for the light long distance users than for the heavy long distance users. Since Kuritsky et al. report results for a difference in usage stimulation between two campaigns, it is of course still possible that the absolute usage stimulation due to each campaign is such that heavy users respond proportionately more than light users. Viewed together, our findings and those from the AT&T study strongly suggest that the magnitude of the differential response between light and heavy users depends on the advertising copy.

An alternate strategy that has been used to stimulate demand is to promote specific applications. For example, some years ago New York Telephone embarked on an advertising campaign whose messages were driven by specific telephone applications (a “thank-you” call, a “reminder” call, a “check-before-you-go” call). Wansink and Ray (1993), in a laboratory experiment, find that primary demand stimulation may indeed be achieved by messages that suggest new applications. An open question in this regard is whether such messages would stimulate light users or heavy users.

With respect to our finding that customers respond in their calling frequency rather than their calling duration, it may be that advertising messages that stress the benefits (emotional, functional) from telephone calling are more likely to generate a response via an additional call rather than additional conversation time. On the other hand, advertising messages that include price information (economic benefits) may stimulate (or inhibit) calling frequency and calling duration. Therefore, we acknowledge again that the copy that is used may be very relevant in determining the nature of the demand stimulation.

5.4. Exploratory Data Analysis and Response Modeling

In addition to the substantive results, we also would like to reflect upon the methods that we used. We believe we have demonstrated the value of exploratory analysis of the data in model specification and in model refinement.

Our methods combine relatively simple graphical displays, such as scatter plots and Q-Q plots, with an ensemble of nonparametric and model-based analytical techniques.
Graphical displays of the data, or data along with model-estimated values are used to advantage in several empirical papers in marketing, for example Little (1979), Guadagni and Little (1983), Abraham and Lodish (1987), and Eastlack and Rao (1989). The analytical techniques include data transformations, robust non-linear smoothers like LOWESS, and resampling methods like the Bootstrap. These methods, increasingly made more effective and more accessible by the availability of high speed computing, comprise a standard subset of what is often loosely referred to as exploratory data analysis. Although the theory of these techniques is complex, given the appropriate computing environment, they are easy to apply and interpret. Since several contemporary statistical computing platforms support these techniques they are now becoming popular with statistical practitioners in many disciplines.

Exploratory data analysis is particularly useful in the absence of well tested and generally accepted theories that can guide the specification of marketing response models. At one level, the direction of exploration can take account of prior knowledge. For example, in examining the response measure we looked naturally for the relation to the pretrial usage anticipating a relation. At another level, even having the notion of a usage-dependent response, we needed the detailed functional form which we obtained by visual inspection of the exploratory smoothing in Figure 4. In addition to the benefits evident in the formulation of models, exploratory data analysis also aids in their calibration, and in diagnosis for model improvement. Used cautiously, with the appropriate interplay of judgment and formal inference, exploratory techniques are extremely effective in guiding the researcher along a more fruitful path among the myriad alternatives that exist in the analysis of any complex problem.

6. Conclusion

In addressing our primary substantive objective, we find with some confidence that advertising did stimulate telephone usage in our study and that most of the response came from heavy users. While we do not claim that this result by itself is generally applicable, it adds to the growing evidence of the interaction between response and prior usage. In addition, we raise and discuss fundamental issues concerning the nature of advertising response and its implications, in particular for telephone companies, and more generally for companies in similar environments such as financial services. Through our analytical process we have demonstrated how exploratory data analysis precedes and guides formal modeling, providing substantive insights along the way. From a methodological perspective, we combine judgment and statistical inference in a highly interactive mix of exploratory data analysis with formal modeling. We believe that our methodology is indeed transferable to other settings to measure change in highly variable consumption or purchase patterns. We hope that our work will stimulate others on these substantive and methodological issues.3

Acknowledgments. The authors thank the advertising director of the telephone company where the experiment took place for sharing background information and the data, as well as insights into the challenges of implementing such an experiment. The results and conclusions of this report are based on the authors' research and do not necessarily reflect the views of the telephone company. Elsa Ancmon, Thomas Ball, Jon Kettenring, Don Morrison, Ambar Rao, Vithala Rao, David Schmittlein and Bob Shoemaker provided helpful comments on an earlier version of this paper. The authors also thank the Editor, the Area Editors and two anonymous reviewers for their many suggestions that contributed to this version of the paper.

3 This paper was received July 16, 1993, and has been with the authors 1 month for 1 revision. Processed by Scott A. Neslin, Area Editor.

Appendix A: Determining the Weights for Correcting Heteroscedasticity

Casual inspection of Figure 4 and the initial estimation of the full model indicates the presence of heteroscedasticity in the relation of the untransformed or transformed trial/pretrial ratio to the pretrial usage level.
Correcting for any heteroscedasticity will provide a better (less biased) estimator of the variance on which our final inference will depend. The power-transformed ratio $r(m)$ for a power-transformed pretrial mean of $m$ is a piecewise linear function of $m$ plus a residual $\epsilon(m)$, which has a normal distribution with mean 0 and a standard deviation of $\sigma(m)$. It can be shown that $|\epsilon(m)|^{1/4}$ has a roughly normal distribution with expectation $k \times E(|X|^{1/4})$, where $k = E(|X|^{1/4})$ is a constant, and $X$ is a random variable with standard normal distribution.

Figure A.1(A) shows the fourth roots of the absolute value of the sample residuals plotted against $m$. Also included is the LOWESS estimate of $E(|\epsilon(m)|^{1/4}) = k\hat{\sigma}(m)^{1/4}$. Evidently $\hat{\sigma}(m)$ is not monotonic in $m$ and attains its smallest value around $m = 2.8$. The estimated weights $w(m)$ are calculated as $w(m) = 1/\hat{\sigma}^2$, and are normalized such that the largest weight equals 1. The function $w(m)$ is shown in Figure A.1(B).

Appendix B: Confidence Intervals Using the Bootstrap

For weighted least squares regression the standard least squares formulae for the standard errors yield exact results only if the correct weights are known. If the weights are estimated from the same data that are used in estimating the regression, these least squares formulae provide standard errors that lead to confidence intervals whose actual coverage probability is only an approximation of the nominal one. Simple closed form confidence intervals with a given coverage probability do not exist. To evaluate the quality of the least squares approximation one may compare the standard errors calculated using the least squares formula with the standard errors based on alternative methods. One such method is the bootstrap; see Efron and Tibshirani (1983).

Consistent with the original sampling plan, each bootstrap replication is implemented as follows: A simple random sample with replacement, called the bootstrap sample, is taken from the set of 735 customers. Note
that some customer households are included in the bootstrap sample more than once, while others are not included at all. Given a specific bootstrap sample, we compute afresh their pretrial averages, response ratios, the regression weights (according to the calculation procedure in Appendix A), and the coefficients for the final model using weighted least squares regression. This process is repeated 500 times, yielding 500 quadruples of bootstrap coefficients. This sample of bootstrap coefficients can be directly used to assess the uncertainty in the estimates of the regression coefficients.

Figure B.1 shows histograms of the bootstrap coefficients for two of the parameters. For the test-control difference of the slope of the right leg, only three of the 500 bootstrap coefficients fall below 0, allowing the inference that the slope difference is indeed positive.

The bootstrap distribution can also be used in a different fashion. The sample standard deviation of the bootstrap coefficients is an estimate of the standard deviation of the regression coefficient estimator. A comparison of this estimate with the standard errors of the weighted least squares estimation is shown below. The small differences suggest that the weighted least squares standard errors are sufficiently accurate.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>WLS</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_c$</td>
<td>-0.0043</td>
<td>0.0044</td>
</tr>
<tr>
<td>$\beta_{cl}$</td>
<td>0.0173</td>
<td>0.0173</td>
</tr>
<tr>
<td>$\beta_{ct}$</td>
<td>0.0107</td>
<td>0.0126</td>
</tr>
<tr>
<td>$\beta_{ct} - \beta_{cl}$</td>
<td>0.0116</td>
<td>0.0122</td>
</tr>
</tbody>
</table>

References


Statistical Sciences Inc. (1990), *Splus Software for Statistical Analysis*, Seattle, WA.


