DYNAMICS OF RETAIL ADVERTISING: EVIDENCE FROM A FIELD EXPERIMENT

DUNCAN SIMESTER, YU (JEFFREY) HU, ERIK BRYNJOLFSSON and ERIC T. ANDERSON*

We use a controlled field experiment to investigate the dynamic effects of retail advertising. The experimental design overcomes limitations hindering previous investigations of this issue. Our study uncovers dynamic advertising effects that have not been considered in previous literature. We find that current advertising does affect future sales, but surprisingly, the effect is not always positive; for the firm's best customers, the long-run outcome may be negative. This finding reflects two competing effects: brand switching and intertemporal substitution. We also find evidence of cross-channel substitution, with the firm's best customers switching demand to the ordering channel that corresponds to the advertising. (*JEL* L2, L81, M3)

A firm's current advertising is generally associated with an increase in its sales, but this effect is generally short-lived.

—*Bagwell* (2005, p. 30).

I. INTRODUCTION

It is accepted that advertising can have a short-run effect on sales. However, our understanding of the dynamic effects of advertising is incomplete. For example, a current advertisement for the retailer Land's End can increase immediate sales, but how will the advertisement affect subsequent consumer demand? If there are long-run effects, how will they vary among consumers? In this article, we empirically investigate the dynamic effects of advertising by analyzing a controlled field experiment for a national, durable goods retailer.

Previous empirical studies have been plagued by both endogeneity and measurement issues, as Bagwell (2005) suggests in his review of the literature. Advertising decisions are endogenous, and so effects attributed to variations in advertising expenditure may actually reflect factors that led to the variation in expenditure. The measurement problems reflect the dynamic nature of advertising outcomes. If all the outcomes were immediate, then measurement would be straightforward. However, measuring long-run outcomes requires a control for the potential confound introduced by intervening events.

In the past, empirical studies have addressed these issues by introducing more sophisticated econometric models (Ackerberg 2001, 2003). In this article, we overcome endogeneity and measurement problems using an

*We thank seminar participants at Georgia Institute of Technology, MIT, Northwestern University, Purdue University, University of Connecticut, University of Maryland, University of Pennsylvania, University of Southern California, 2004 Workshop on Information System and Economics, 2005 Symposium on Electronic Commerce Research, 2005 Fifth Annual INFORMS Revenue Management and Pricing Conference, 2006 Economic Science Association North American Meeting, and 2006 National Bureau of Economic Research Industrial Organization Winter Meeting. Generous funding was provided by MIT Center for Digital Business.

Simester: NTU Professor of Management Science, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142. Phone 1-617-258-0679, Fax 617-258-7597, E-mail simester@mit.edu

Hu: Assistant Professor, Krannert School of Management, Purdue University, West Lafayette, IN 47907. Phone 1-765-494-7907, Fax 765-494-9658, E-mail yuhu@purdue.edu

Brynjolfsson: Schussel Family Professor and Director of MIT Center for Digital Business, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142. Phone 1-617-253-4319, Fax 617-258-7579, E-mail erikb@mit.edu

Anderson: Hartmarx Research Associate Professor of Marketing, Kellogg School of Management, Northwestern University, Evanston, IL 60208. Phone 1-847-467-6482, Fax 847-491-2498, E-mail eric-anderson@kellogg.northwestern.edu

ABBREVIATION

RFM: Recency, Frequency, and Monetary Value

Economic Inquiry
(ISSN 0095-2583)
Vol. 47, No. 3, July 2009, 482–499

doi:10.1111/j.1465-7295.2008.00161.x
Online Early publication July 9, 2008

alternative approach. We conduct a field experiment in which we experimentally vary advertising levels for two randomly selected samples of customers.[1] The field setting ensures that consumers engage in actual market transactions and interact with the retailer in a natural manner. The experimental manipulation introduces exogenous variation in advertising, which overcomes endogeneity concerns. Moreover, because customers in each experimental condition are exposed to the same intervening events, we can overcome this confound by comparing demand across the two experimental conditions. This allows us to draw clear causal inferences about the effects of advertising.

There have been many empirical studies of advertising in the consumer packaged goods industry (e.g., Ackerberg 2001, 2003; Deighton, Henderson, and Neslin 1994; Tellis 1988) but few studies in durable goods markets. Our study confirms that advertising works differently in a durable goods market. In particular, we provide evidence of two competing effects. First, we show that advertising can affect purchase timing. It is well known that price reductions can cause intertemporal substitution and we show that advertising by a durable good firm may lead to similar effects. An implication of this finding is that advertising may have an initial positive effect on demand followed by a later negative impact. Second, we show that advertising can cause a significant increase in future demand. This second effect is consistent with a long-run goodwill effect and can lead to increases in both short-run and long-run demands.

We show that the magnitude of these effects varies systematically across consumers. Intertemporal substitution is the dominant effect among the firm's "Best" customers who had historically placed a large number of orders with the firm. For these customers, the short-run increase in demand is almost entirely offset by a *reduction* in future demand. This is a robust result that survives a series of validity checks. We interpret the result as evidence of intertemporal substitution—an effect that has not been previously recognized in the advertising literature.

We also show that advertising causes cross-channel substitution. Customers in this study place their orders through two types of channels: the catalog channel (mail and telephone sales) and the Internet channel. We find that catalog advertising leads to a short-run increase in catalog orders for all customers. However, the impact on Internet orders varies. Among the firm's Best customers, there is an increase in catalog demand and a reduction in Internet demand, while for other customers, catalog advertising leads to increased demand in both channels.

Overall the results indicate that advertising can do little to make a firm's Best customers any better. For these customers, any short-run increases in demand reflect substitution either from different channels or from future demand. In contrast, for other customers, advertising can lead to a long-run shift in the demand function. For these customers, the increase in short-run demand is complemented by higher demand in other channels and higher demand in future periods.

The field experiment was conducted in a retail advertising setting by varying the number of direct mail catalogs sent to customers of a women's clothing retailer. Retail catalogs clearly contain elements that are generally accepted as advertising. One might also conjecture that retail catalogs have other effects that are not typically considered advertising. Is a Land's End catalog the same type of advertising as a Coca-Cola television commercial? We point to two distinctive features of our study. First, there are clear differences between retail and manufacturer advertising. Second, advertising effects may differ by media type. We discuss each of these next.

II. DISTINGUISHING CATALOG ADVERTISING FROM OTHER FORMS OF ADVERTISING

While other articles have analyzed retail advertising (Milyo and Waldfogel 1999), the distinction between manufacturer and retail advertising has received little attention in the economics literature. In contrast, these differences are well documented in the advertising literature. For example, Wells, Moriarity, and Burnett (2006) observe that retail advertising focuses on influencing *where* customers purchase rather than simply *what* they purchase. The content of retail advertising typically provides information about multiple items and includes specific details about how to make a purchase. This includes information

1. The study is a "natural field experiment" under Harrison and List's (2004) nomenclature.

about shopping hours, acceptable payment methods, and ordering options together with directions to the retailer's Internet or physical stores. On the other hand, manufacturer advertising rarely contains information about multiple products or details about shopping hours or payment methods (though it is not unusual to identify alternative retailers).

The retail catalogs involved in this study describe the retailer's ordering procedures, hours of operation, warranties, and payment methods. The majority of products in the catalog are private label and carry the retailer's brand name, and so the images, copy, and catalog design emphasize the retailer's brand rather than the manufacturers of the various products. The situation is similar to Tiffany & Co.'s powder blue catalogs, which reinforce Tiffany's brand image and provide detailed information about its products. This contrasts with a manufacturer advertisement (e.g., for Sony or Coca-Cola), which emphasizes product characteristics and the manufacturer's brand image.

There are also clear differences across advertising media. Catalog advertising is a form of print media, and a characteristic of this media is that exposure is controlled in part by customers. Print media may also be easily stored, so that customers need not rely on memory to retrieve the advertising content. For example, a customer may be exposed to a magazine advertisement, store the magazine, and later retrieve the advertisement to find a phone number, Web address, or product information. This contrasts with broadcast media, such as radio and television, where advertising is consumed in real time.

Print media is recognized by the advertising industry as the dominant form of advertising. In 2003, a total of $245 billion was spent on advertising in the United States with direct mail ($48 billion) and newspapers ($45 billion) representing the two largest categories. Despite the level of expenditure, direct mail and other forms of print media have gone largely unnoticed in the economics literature.[2] The lack of research is surprising not just because of the economic importance. As we will discuss, direct mail advertising offers important measurement advantages over both manufacturer advertising and other advertising media. In this study, we are able to track the historical and future purchasing behavior of individual customers and form a causal link between the experimental manipulations and the subsequent change in customer behavior.

A. Prior Theoretical Work

Much of the theoretical advertising literature has focused on distinguishing whether advertising serves a persuasive or informative role. Under the persuasive view, advertising enters customers' utilities for different products (Becker and Murphy 1993; Comanor and Wilson 1967, 1974; Kaldor 1950). This leads to an outward shift in the demand function, which has led to claims that advertising may serve an important anticompetitive role. Under the informative view, advertising increases the information that customers have about the available alternatives (Kihlstrom and Riordan 1984; Milgrom and Roberts 1986; Stigler 1961).[3] The persuasive and informative views of the role of advertising are both consistent with advertising positively impacting demand in future periods. Yet, it is also possible that the long-run impact of advertising is negative. When making purchasing decisions, customers generally have the alternatives of purchasing competing brands, purchasing from different retailers, or even delaying in the hope of future discounts or product improvements. If advertising makes an immediate purchase of the focal brand more attractive, it implicitly reduces the share of customers who will choose one of these alternatives. The outcome is potentially less demand for competing brands, less demand for competing retailers, and/or less demand in future periods. Of these alternative outcomes, the impact on competing brands (sometimes termed the "combative" role of advertising) has received the most interest. Borden (1942) distinguished between the "primary" and the "selective" effects of advertising: the primary effect describes category-level demand expansion, while the selective effect describes substitution between competing brands. More recently, the distinction between advertising's primary and selective effects has served as a central focus of debate in the

2. This contrasts with Internet advertising, which despite its recency and relatively small size has attracted considerable attention from economists.

3. See also Telser (1964), Nelson (1970, 1974), Schmalensee (1978), and Grossman and Shapiro (1984).

tobacco industry (e.g., Roberts and Samuelson 1988; Seldon and Doroodian 1989). The industry has sought to ward off proposed regulation limiting tobacco advertising by arguing that advertising serves primarily a selective role, allowing companies to attract share from their competitors without expanding total industry demand. In contrast, antismoking advocates have argued that tobacco advertising also has an impact on primary demand, contributing to an expansion in total tobacco consumption.

Substitution between brands is analogous to substitution across time. In many product categories, purchasing a competing brand and purchasing in future periods both represent alternatives to making an immediate purchase of the focal brand. Although the possibility of intertemporal substitution has received relatively little attention in the advertising literature, it has received considerable attention in the pricing literature. There is well-documented evidence that price discounts can lead to both brand substitution and intertemporal substitution. As a result, following a price promotion, there is often evidence of a "postpromotion dip" in sales, as customers consume products purchased during the discount period (Blattberg and Neslin 1990, p. 358; Hendel and Nevo 2003).[4] Interestingly, there is also evidence that this intertemporal effect varies across customers (Anderson and Simester 2004). The negative long-run effect of a price promotion appears to be most pronounced for customers who have the most experience with the brand.

We conclude that there is theoretical support for advertising having both a positive impact and a negative impact on future demand. If advertising increases customers' expected utility through persuasion or information and this increase is enduring, the impact on future demand will tend to be positive. On the other hand, if advertising accelerates demand, intertemporal substitution may lead to a negative impact on future demand.

B. Prior Empirical Evidence

There is some evidence of a positive long-run relationship between advertising and sales. Yet, many studies report either that no long-run impact or the impact is short lived (Bagwell 2005). There are no studies reporting a negative relationship between advertising and future demand. However, as we recognized, this empirical work has been confronted by important challenges. Early research on advertising was often limited to aggregate brand or category-level data in which researchers investigated the relationship between current advertising and lagged effects on sales. Because the sign of the effect could theoretically vary for different subsets of consumers, aggregate data may not detect an effect even when it is present. These studies also suffered from important limitations due to both the possibility of intervening events and the potential endogeneity of advertising decisions (Lambin 1976; Schmalensee 1972). More recently, the development of household-level panel data sets has made it possible to estimate demand at the individual or household level. Together with methodological developments, these new data sets offer the opportunity to address endogeneity through advanced econometric controls (e.g., Ackerberg 2001, 2003; Erdem and Keane 1996).

In contrast to exploiting more sophisticated econometric methods, our approach is to improve the measurement of the relationship between advertising and sales. Random assignment of customers to high advertising and low advertising groups introduces an external control to the data collection process, which helps prevent the introduction of confounds. This contrasts with previous studies in which researchers have had to accept the presence of confounds in their data and instead sought to provide internal controls for these confounds in their analyses. The experimental approach also offers another advantage: the results are easily analyzed and interpreted. The experimental design yields a simple comparison between groups of customers who experience one advertising treatment and equivalent control groups who experience a different level of advertising. We directly measure (and interpret) the difference in customers' long-run demand.

This is not the first field experiment designed to investigate the impact of advertising. Managerial studies using proprietary split-sample cable television experiments have previously been used in the consumer packaged goods industry. Unfortunately, academic descriptions of these findings are necessarily limited by the proprietary nature of the data, models, and parameter estimates (e.g., Aaker

4. See also Hendel and Nevo (2002, 2005).

and Carman 1982; Lodish et al. 1995a, 1995b). Moreover, the results of these studies are mixed, which may reflect a lack of statistical power.[5] There have been at least two academic studies that use experiments to investigate how advertising influences prices and price elasticities. Krishnamurthi and Raj (1985) report the findings from a split-sample cable television experiment and conclude that advertising is capable of reducing consumer price elasticities. More recently, Milyo and Waldfogel (1999) use a natural experiment to study the effect of retail advertising on prices. They find that advertising does tend to lower the retail prices of advertised products but has little effect on the prices of unadvertised products.[6] There have also been a small number of studies investigating how varying the advertising message can influence the response rate. In a recent example, Bertrand et al. (2006) use a randomized direct mail experiment to measure how changing features of an advertisement for consumer credit affected the response rate.

C. Structure of the Article

The article proceeds in Section III with a simple model illustrating the intuition that current advertising may lead to a positive impact or a negative impact on future demand. We then provide an overview of the study design in Section IV before presenting the results in Section V. The results section begins with a review of the short-run impact followed by the long-run and cross-channel outcomes. We then investigate alternative explanations for the findings by comparing the heterogeneity in the results across different customer segments. The article concludes in Section VI with a review of the findings and implications.

III. POSITIVE AND NEGATIVE LONG-RUN OUTCOMES

To help understand why current advertising may lead to a positive impact or negative impact on future demand, we highlight two opposing retail advertising effects: brand switching and intertemporal substitution.

We consider a focal firm that produces a different product in two periods. Competing firms offer imperfect substitutes, which we collectively describe as a single outside option. Customers purchase q_t units of the focal firm's products in each period t and purchase $\bar{q}$ units of the outside option. Consumers have utility $U(q_1, q_2, \bar{q}|v_1, v_2, \bar{v})$, where v_t and $\bar{v}$ are product preference parameters, and the marginal utility of each product is increasing in its preference parameter: $\mathrm{d}^2U/\mathrm{d}q_t\,\mathrm{d}v_t > 0$.

Customers face a budget constraint: $Y = p_1q_1 + p_2q_2 + \bar{p}\bar{q}$, where p_t and $\bar{p}$ are product prices. As our focus is on advertising effects, prices do not play an important part in the model. Therefore, we scale all prices to one, which allows us to interpret the budget constraint as a consumption constraint. For example, in the apparel setting described by our empirical findings, consumers may have a limit on the number of items that they can purchase in a season. This constraint may result from physical storage constraints (available wardrobe space) or limited consumption opportunities (customers with too many clothes cannot wear them all).[7]

The v_t terms in the utility function are preference parameters that are influenced by advertising. We make the natural assumptions that v_t is increasing in both current and prior period advertising and that carryover to future periods decays over time: $\mathrm{d}v_t/\mathrm{d}a_t > \mathrm{d}v_t/\mathrm{d}a_{t-j} > 0$ for all $j > 0$, while $\mathrm{d}v_t/\mathrm{d}a_{t+k} = 0$ for all $k > 0$. Intuitively, we could interpret advertising as having firm-specific and time-specific effects: while the firm effects endure, the time-specific effects decay. We also assume that advertising by the focal firm does not directly affect preferences for the competing product: $\mathrm{d}\bar{v}/\mathrm{d}a_t = 0$. While the relationship between advertising and preferences for the focal brand is positive, we do not seek to distinguish between the *information* and the *persuasion* interpretations proposed in the literature.

Customers have a discount factor equal to 1 and select the quantity of goods $(q_1^*, q_2^*, \bar{q}^*)$ that maximizes utility for both periods subject to their budget constraint. We will assume that customers are forward looking and that they anticipate the effects of advertising may decay over time. While it is convenient to assume that customer forecasts are accurate, our arguments

5. These three articles do not report sample sizes or the estimation models for individual studies.

6. For other examples of natural experiments, see Benham (1972) and Ippolito and Mathios (1990).

7. We thank Steven Tadelis for this observation.

do not critically depend on their accuracy. We merely require that customers do not (erroneously) anticipate the opposite outcome: that the effects of advertising will grow over time.

Since current advertising increases current utility, current advertising causes an increase in current demand: $\mathrm{d}q_1^*/\mathrm{d}a_1 > 0$. However, the impact of current advertising on future demand ($\mathrm{d}q_2^*/\mathrm{d}a_1$) is ambiguous and reflects a trade-off between brand switching and intertemporal substitution. Because the effects of advertising persist, advertising in Period 1 makes the firm more attractive in Period 2 compared to the alternative of purchasing from one of the competitors. This leads to brand switching in which customers shift demand from the outside option to the focal firm. On the other hand, because the effects of advertising decay over time, purchasing in the first period is relatively more attractive to purchasing in the second period. This leads to intertemporal substitution in which second period demand is shifted to the first period.

The extent of brand switching and intertemporal substitution depends upon customers' preferences for the firm. This is best illustrated by considering customers who have such strong preferences for the firm that they only buy the focal firm's products. Because these customers do not buy from the competitor, advertising cannot lead to brand switching. This limits the outcome to intertemporal substitution, and so current advertising causes a reduction in future demand among the firm's Best customers. Intuitively, advertising to the firm's Best customers cannot make them any better. If they are already buying all their products from the firm, advertising can only shift demand between periods. For customers with weaker preferences, brand switching is relevant and so demand for the focal firm may also increase in the second period. Thus, when customers' preferences for the firm are weaker, we may observe a favorable long-run advertising response. In the Appendix, we provide a derivation of these predictions for a specific utility function.

We conclude that we expect the long-run effect of advertising to vary systematically across customers. The study described in the next section provides an opportunity to test these predictions. Specifically, we will evaluate whether the long-run response to advertising is moderated by the strength of customers' prior preferences for the firm. We measure these prior preferences using customers' historical purchasing behavior.

IV. STUDY DESIGN

The study was conducted with a medium-sized company that sells women's clothing in the moderate price range.[8] All the products carry the company's private label brand and are sold exclusively through the company's own catalogs, Internet Web site, and retail stores. The study involved a total of 20,000 customers who had previously made a purchase from the company through the catalog channel (mail or telephone) or Internet channel. To explore the effects of heterogeneity in the sample, the company initially identified two distinct samples of customers. The first sample of 10,000 customers, which we denote the Best customers, were all customers who had made relatively frequent and recent purchases from the company. In particular, these were the customers whom the company's own statistical models suggested would be most likely to purchase if mailed a catalog.[9] The "Other" sample of 10,000 customers comprise customers who the company's statistical model predicted had an average probability of responding if mailed a catalog.

Within the Best and Other customer groups, customers were randomly assigned into equal-sized high advertising and low advertising conditions. This yielded a total of four different customer samples (Table 1). In each case, the final sample sizes were slightly smaller than 5,000. The reason for this is rather technical but does not affect the interpretation of the study.[10]

The experimental manipulation occurred over an 8-mo period. During this period, all

8. The company asked to remain anonymous.

9. Although the details of the company's statistical models are proprietary and were not made available to the research team, we found that the recency and frequency of prior purchases accurately distinguish these customers.

10. Because customers rarely have their unique customer identification numbers available when they call to place an order, individual customers sometimes end up with more than one account number. Each month, the company uses various methods to identify these duplicate account numbers and consolidate them back to a single account number. The reduction in the sample sizes reflects the deletion of duplicate account numbers. Fortunately, this process is identical for the treatment and control samples and so cannot explain systematic differences between them.

TABLE 1
Sample Sizes

	Low Advertising Sample	High Advertising Sample
"Best" customers	4,921	4,904
"Other" customers	4,790	4,758

TABLE 2
Mailing Dates in 2002 by Experimental Condition

	Low Advertising	High Advertising
Catalog 1		
Mailing Date 1	January 11	January 11
Mailing Date 2	February 22	February 8
Catalog 2		
Mailing Date 1	February 1	January 25
Mailing Date 2		February 22
Catalog 3		
Mailing Date 1	March 15	March 8
Mailing Date 2	April 26	April 5
Catalog 4		
Mailing Date 1	April 5	March 22
Mailing Date 2		May 3
Catalog 5		
Mailing Date 1	May 17	April 19
Mailing Date 2		May 17
Catalog 6		
Mailing Date 1	June 7	June 7
Mailing Date 2	June 28	June 28
Catalog 7		
Mailing Date 1	July 26	July 26
Mailing Date 2	September 6	August 23
Mailing Date 3		September 20
Catalog 8		
Mailing Date 1	August 9	August 9
Mailing Date 2		September 6

the customers in the high advertising sample received a total of 17 catalogs, while customers in the low advertising sample received just 12 catalogs.

We use the "high advertising" and "low advertising" labels merely for expositional convenience. In the absence of the study, the actual number of catalogs mailed would have varied across the customers. For example, a comparison of the mailing strategies for the same customers across the same period in the previous year reveals that the *Other* customers received an average of 12.2 catalogs, while the *Best* customers received an average of 14.9. For both samples, the maximum and minimum mailing frequencies were 19 and 1, respectively. We caution that because neither of our experimental conditions represents what the firm would have done in the absence of the test, we cannot directly evaluate the optimality of this firm's existing policy. However, we will later use the findings to describe how a myopic approach to making mailing decisions would lead to a suboptimal outcome.

The additional catalogs sent to the high advertising sample were simply additional copies of catalogs that all customers received. This ensured that the experimental manipulation only affected the frequency of advertising and not which products were available or features specific to the design of the catalogs. Sending multiple copies of the same catalog to the same customer is a common practice in the catalog industry. Designing new catalogs is expensive and so to reduce their design costs firms often resend the same catalog 2–4 wk after the first mailing.

In Table 2, we summarize the mailing schedule in each condition for the eight different catalogs used in the test. The specific timing of each mailing was determined by the company's circulation managers who were instructed to optimize the overall (short run) response given the exogenous decision to mail 12 times in the low condition and 17 times in the high condition. It is possible that varying the timings would lead to differences in the long-run results. It is also difficult to speculate how the findings would be affected if we had chosen different mailing frequencies. Following the experimental manipulations, the company returned to using its standard mailing policies and made no distinction between customers in the two conditions.

All eight catalogs were regularly priced catalogs and contained a similar number of items. Because the mailing dates coincide with different fashion seasons, the main difference between the eight catalogs is the product selection. In Table 3, we summarize the seasons for which each catalog was targeted together with the average price paid for items purchased from each catalog. We see that the average price was almost identical, except for the last two catalogs, where the focus on fall clothing

TABLE 3
Catalog Content

	Season	Average Price Paid
Catalog 1	Spring	\$67.76
Catalog 2	Spring	\$67.13
Catalog 3	Summer	\$68.34
Catalog 4	Summer	\$68.08
Catalog 5	Midsummer	\$65.08
Catalog 6	Early fall	\$65.03
Catalog 7	Fall	\$85.42
Catalog 8	Fall	\$78.79

led to an increase in the average price (fall clothing tends to be more expensive).

Because the first catalog was mailed to both samples on the same day, the date of the first manipulation was actually January 25, 2002 (when only customers in the high advertising group were sent Catalog 2). The last date on which the mailing policies were different for the two samples was September 20, 2002. We received data describing the number of items purchased by customers before, during, and after the experimental manipulations. In particular, we received a record of all transactions made from January 1, 1988, until almost 19 mo after the start of the first manipulation (August 13, 2003). To simplify the analysis and discussion of the results, it is helpful to define three periods:

(1) The "pretest" period: from January 1, 1988, through January 24, 2002.
(2) The "test" period: from January 25, 2002, through December 31, 2002.
(3) The "posttest" period: from January 1, 2003, through August 13, 2003.

Notice that the test period extends for 103 d beyond the date of the last manipulation: September 20, 2002, through December 31, 2002. This was designed to capture orders from catalogs mailed toward the end of the test period. The company estimated that more than 99% of the immediate demand from catalogs mailed in September would have occurred by December 31. This is also consistent with the industry-wide response curve reported by the Direct Marketing Association (2003). We later vary the length of the posttest period to investigate how it affects the results.

We caution that the transaction data only involves customers' purchases through the company's Internet Web site or its catalog channel (mail and telephone orders). This represents approximately 65% of the company's total sales, with the remaining transactions occurring at its retail stores. We do not have a record of purchases made by these customers in the company's retail stores because at the time of the study, the company was unable to adequately identify customers purchasing in its stores. We will later investigate how this omission may have affected the results by restricting attention to customers who live a long way from the company's stores.

The historical purchasing results provide a means of checking whether the assignment of customers to the high advertising and low advertising conditions was truly random. In particular, in Table 4, we compare the average *recency*, *frequency*, and *monetary* value (RFM) of customers' purchases during the pretest period.[11] We also include a measure of the time required to drive (in hundreds of minutes) between the customer's mailing address and the nearest store operated by the firm. We obtained this measure by querying a driving time database using the street address for each customer and each store location. We then identified the nearest store using the minimum driving time. As noted above, we will later use this measure to control for the possibility of cannibalization from store demand.

If the random assignment was truly random, we should not observe any systematic differences in these historical measures between the high advertising and low advertising samples. The findings reveal no significant differences in the historical demand in either the Best or the Other customer samples.

V. RESULTS

A. Does Current Advertising Impact Short-Run Demand?

In Table 5, we summarize demand in the high advertising and low advertising conditions during the test period and report both univariate and multivariate comparisons. The univariate analysis is simply the average

11. "Recency" is measured as the number of days (in hundreds) since a customer's last purchase. "Frequency" measures the number of items that customers previously purchased. "Monetary value" measures the average price (in dollars) of the items ordered by each customer.

TABLE 4
Check on Randomization Process Historical Purchases during the Pretest Period

	Low Advertising Condition	High Advertising Condition	p Value
Best customers			
Recency	1.43 (0.02)	1.43 (0.01)	.72
Frequency	40.38 (0.45)	40.75 (0.51)	.59
Monetary value	61.11 (0.19)	61.22 (0.19)	.69
Driving time	1.66 (0.02)	1.68 (0.02)	.46
Other customers			
Recency	4.67 (0.06)	4.76 (0.06)	.30
Frequency	10.56 (0.20)	10.62 (0.21)	.85
Monetary value	63.85 (0.29)	64.18 (0.33)	.50
Driving time	1.74 (0.02)	1.76 (0.02)	.81

Notes: The table reports the average values of each variable for each subsample. Standard errors are given in parentheses. The p values denote the probability that the difference between the high and low advertising averages will be larger than the observed difference (under the null hypothesis that the true averages are identical).

number of items purchased by customers in each sample. The multivariate analysis uses customers' historical (pretest) purchases to control for individual customer characteristics. The unit of observation is a customer (denoted by subscript i), and the dependent measure is the number of items purchased during the test period (Q_i). Because Q_i is a "count" measure, the multivariate analysis uses Poisson regression. We model the purchase rate as $\ln \lambda_{it} = \beta X_{it}$ and use the following specification for the independent variables:

$$\begin{aligned}\ln(\lambda_i) = \beta_0 + \beta_1 \text{high advertising}_i \\ + \beta_2 \log(\text{recency}_i) + \beta_3 \log(\text{frequency}_i) \\ + \beta_4 \log(\text{monetary value}_i) \\ + \beta_5 \log(\text{driving time}_i).\end{aligned} \quad (1)$$

The variable of interest is *high advertising*, which is a dummy variable identifying whether customer i was in the high advertising condition. Under this specification, β_1 measures the percentage change in short-run demand between customers in the high advertising condition compared to those in the low advertising condition. The specification also preserves the benefits of the randomized experimental design, providing an explicit control for intervening factors such as competitors' actions and macroeconomics.

The control variables include the RFM of customers' prior purchases. Recall that we earlier used these variables to check the validity of the randomization procedures (Table 4). They are well-established metrics for segmenting customers in this industry and provide natural candidates to control for differences in customers' historical purchasing patterns. For completeness, we also include the driving time variable, measuring the time required to drive between the customers' mailing address and the nearest store operated by the firm. We estimate separate models for the Best and Other customers and report the results for both samples in Table 5. The small difference in the univariate and multivariate sample sizes reflects the absence of driving time data for a handful of customers. The omission of these customers has essentially no impact on the results.

The findings reveal that the additional advertising received by the high advertising sample led to a significant short-run increase in demand for both the Best and the Other customers. The demand increase was approximately

TABLE 5
Units Ordered during the Test Period

	Other Customers	Best Customers
Univariate analysis		
Low advertising condition	1.08 (0.04)	3.63 (0.08)
High advertising condition	1.24 (0.05)	3.86 (0.09)
Difference	0.16* (0.07)	0.23* (0.12)
Sample size	9,548	9,825
Multivariate analysis		
High advertising condition	0.138** (0.019)	0.053** (0.011)
Recency	−0.276** (0.006)	−0.132** (0.004)
Frequency	0.485** (0.010)	0.747** (0.008)
Monetary value	0.420** (0.029)	0.843** (0.026)
Driving time	−0.014 (0.009)	0.006 (0.005)
Intercept	−1.188** (0.134)	−4.309** (0.118)
Log likelihood	−19,160	−33,919
Sample size	9,458	9,761

Notes: The univariate analysis reports the average number of units purchased during the test period. The multivariate analysis reports the coefficients from Equation (1). Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; ** significantly different from zero, $p < .01$.

5.3% for the Best customers and 13.8% for the Other customers. In percentage terms, the demand increase was significantly larger among the Other customers, but this was calculated over a smaller base. In absolute terms, the effect was not significantly different across the two populations. We conclude that current advertising can cause a significant increase in short-run demand. While these results provide a reassuring manipulation check, they are not the main focus of this article. Instead, we are interested in learning how increasing current advertising affects demand in *future* periods.

B. Does Current Advertising Impact Future Demand?

The long-run impact of the experimental manipulation on posttest demand is summarized in Table 6. For the sake of brevity, we restrict attention to the multivariate analysis and only report the coefficients for the high advertising variable (complete results are provided in Appendix Table A1). As a basis of comparison, we repeat the corresponding coefficients for the test period (Table 5) and also report the coefficients when combining the data from both the test and posttest periods.

The findings reveal a strikingly different picture for the Best and Other customers. Among the Other customers, the increased demand in the high advertising condition persists throughout the posttest period. The effect size decreases from 13.8% in the test period to 10.0% in the posttest period, but this difference is not significant. Among the Best customers, we also see a significant long-run effect. However, the sign of the effect is reversed, with the increase in demand during the test period offset by a significant loss of demand in the posttest period. This pattern is consistent with temporal substitution in which customers shift purchases from the posttest period to the test period.

To our knowledge, this is the first evidence of a significant negative long-run effect attributed to advertising. Similar results have been reported for price promotions, but price variation cannot explain the findings in this study. While we manipulated the frequency of mailings, the prices and other catalog content were held constant.

TABLE 6
Comparison of Test Period, Posttest Period, and Overall Results

	Other Customers	Best Customers
Test period	0.138** (0.019)	0.053** (0.011)
Posttest period	0.100** (0.026)	−0.036** (0.013)
Overall: test and posttest periods	0.125** (0.016)	0.018* (0.008)
Sample sizes	9,458	9,761

Notes: The table reports the *high advertising* variable coefficients when estimating Equation (1) separately on the test period, posttest period, and overall period data sets. Complete findings (including the omitted coefficients) are reported in Table 5 and Appendix Table A1. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

C. Persistence of the Effect

Recall that the posttest period extended from January 1, 2003, through August 13, 2003. It is possible that the adverse outcome persists beyond this period. To investigate this possibility, we divided the posttest period into two equal-sized (112 d) subperiods and repeated the analysis. This allows us to compare the impact of the additional catalog advertising on demand at the start and end of the posttest period. The findings for both subperiods are summarized in Table 7 (complete findings are available in the Appendix).

The negative posttest outcome for the Best customers is concentrated at the start of the

TABLE 7
Comparison of Posttest Results Start and End of the Posttest Period

	Other Customers	Best Customers
Start of posttest period	0.114** (0.038)	−0.096** (0.019)
End of posttest period	0.087* (0.037)	0.021 (0.018)
Complete posttest period	0.100** (0.026)	−0.036** (0.013)
Sample sizes	9,458	9,761

Notes: The table reports the *high advertising* variable coefficients when estimating Equation (1) using data from the start and end of the posttest period. Complete findings (including the omitted coefficients) are reported in Appendix Table A2. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

period. By the end of the period, the effect is no longer apparent. This is consistent with our interpretation that the adverse long-run outcome for these customers reflects intertemporal substitution. In studies of intertemporal substitution in the pricing literature, we see a similar pattern, with the postpromotion dip concentrated immediately after the promotion period and no effect observed on demand in later periods. For the Other customers, the increase in catalog frequency in the high advertising condition leads to a significant increase in demand throughout the posttest period. Although the estimated effect size drops from 11.4% to 8.7% by the end of the period, the difference between the two coefficients is not statistically significant. These findings suggest that the favorable lift in demand for the Other customers may also have extended beyond the posttest period, so that the coefficient reported in Table 6 may underestimate the true long-run effect.

The findings in Tables 6 and 7 also reveal how the findings change as we vary the length of the test and posttest periods. When the demarcation date distinguishing the test and posttest periods is extended beyond December 31, 2002, to also include the start of 2003, we see a drop in the test period effect among the Best customers. The effect is most negative for these customers in the first months of 2003, and so extending the demarcation date into 2003 leads to the inclusion of this negative long-run effect into the test period results. For the Other customers, varying the demarcation date has little impact on the findings.

D. Persistence of the Manipulation

An alternative explanation for the positive long-run effect among the Other customers is that the change in demand during the test period may have affected the mailing policy during the posttest period. Recall that the firm used the same mailing policy for all customers once the experimental manipulation was over. Because this policy tends to mail more frequently to customers with recent purchases, it is possible that customers in the high advertising condition continued to receive more frequent mailings after the experimental manipulation.

Although we do not have data describing the mailing decisions after the experimental treatment, this does not appear to be a complete explanation for the findings. First, an increase in posttest mailings to the Best customers obviously cannot explain the drop in posttest demand among these customers. Among the Other customers, the increase in posttest demand is consistent with more frequent posttest mailings. However, if we restrict attention to customers who made the same number of test period purchases, we can rule out any systematic differences in the posttest mailing policies. Discussions with the firm confirm that its mailing policy only depends on past purchasing and does not consider how many catalogs a customer was previously mailed. Therefore, by focusing on customers who made the same number of test period purchases, we can be confident that there are no differences in the posttest mailing decisions between the two conditions.

The most common outcomes in the test period were that customers purchased zero items or they purchased a single item. When we replicate the posttest analysis using Other customers who made zero test period purchases, the estimated lift in posttest demand is approximately 7%, while for customers with exactly one purchase during the test period, it increases to 19%. Neither of these effects is significantly different from the 9.7% effect reported in Table 6.

Notice that focusing on customers who made the same number of test period purchases introduces a possible selection effect: customers in the high advertising condition are likely to be different than those in the low advertising condition. A natural interpretation is that customers in the high advertising condition are lower probability purchasers (they made the same number of purchases despite receiving more advertising). This works against the observed result, suggesting that replication of the posttest findings in these subsamples occurs *despite* this selection effect. However, it is possible to construct scenarios that reverse the selection effect. For this reason, we interpret this robustness check as indicative but not conclusive.

E. Channel Substitution

Recall that we received demand data for purchases made through both the catalog channel (mail and telephone) and the company's Internet Web site. In the findings reported above, we aggregated test period demand across the catalog and Internet channels. However, by analyzing demand separately for these two channels, we can investigate whether the incremental catalog in the high advertising condition led to substitution from the Internet channel to the catalog channel.

To distinguish the impact of the advertising manipulation on the two ordering channels, we separately calculated the number of items purchased during the test period through the Internet and catalog channels (our data does not distinguish between catalog orders received via mail vs. telephone). We then re-estimated Equation (10) separately using both of these dependent measures. The findings are reported in Table 8. Again, for ease of presentation, we only report the high advertising coefficients (the complete model is reported in Appendix Table A3). The pattern of findings in the cross-channel analysis is analogous to the long-run analysis. The favorable outcome for Other customers extends across both channels. In contrast, among Best customers, the favorable outcome in the catalog channel is offset by a significant reduction in demand over the Internet channel.

We caution that we do not have data describing demand in the company's retail stores. The evidence of channel switching among the Best customers suggests that the increase in catalog advertising may also have switched demand from the retail stores to the catalog channel, at least for customers living close to these stores. In this respect, our measures of the total impact of the test may overlook the change in retail store demand. In our next analysis, we investigate this possibility by restricting attention to customers who live a long distance from any of the firm's stores.

TABLE 8
Comparison of Test Period Results By Channel

	Other Customers	Best Customers
Catalog channel	0.118** (0.020)	0.065** (0.011)
Internet channel	0.281** (0.055)	−0.092* (0.038)
Sample sizes	9,458	9,761

Notes: The table reports the *high advertising* variable coefficients when estimating Equation (1) separately on demand from the catalog channel, demand from the Internet channel, and total demand across both channels. Complete findings (including the omitted coefficients) are reported in Appendix Table A3. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

F. Customers Who Live Far Away from the Firm's Stores

Industry wisdom argues that customers who live more than an hour's driving time from a store are unlikely to purchase from that store. Therefore, to control for any possible impact of the test on store demand, we omitted any customers who live within an hour's driving time of one of the company's stores and then repeated our earlier analysis. The results are summarized in Table 9 (complete results are in Appendix Table A4). For the Best customers, the findings replicate the earlier results: we see an increase in demand during the test period, following by a decrease in demand in the posttest period. For the Other customers, increasing the advertising frequency also led to a significant increase in demand during the test period. However, the change in posttest demand is now smaller and no longer significant.

The smaller posttest effect for Other customers suggests that the posttest outcome for customers living close to the store may in part reflect channel substitution. Because we only measure Internet and catalog demands, if customers who live close to a store switch their posttest demand from the store to the catalog channel, we will observe an

TABLE 9
Comparison of Test Period, Posttest Period, and Overall Results: Customers Living More Than an Hour from a Store

	Other Customers	Best Customers
Test period	0.163** (0.023)	0.082** (0.013)
Posttest period	0.020 (0.032)	−0.048** (0.016)
Overall: test and posttest periods	0.113** (0.019)	0.030** (0.010)
Sample sizes	6,555	6,628

Notes: The table reports the *high advertising* variable coefficients when estimating Equation (1) separately on the test period, posttest period, and overall period data sets. Complete findings (including the omitted coefficients) are reported in Appendix Table A4. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

TABLE 10
Average Profit Earned Per Customer High Advertising vs. Low Advertising

	Low Advertising	High Advertising	Difference
Average profit earned from the Best customers			
Test period, catalog profit	$89.98	$91.56	$1.58
Test period, catalog and Internet profits	$98.74	$100.27	$1.53
Test and posttest periods, catalog and Internet profits	$164.57	$163.84	−$0.73
Sample size	4,921	4,904	
Average profit earned from the Other customers			
Test period, catalog profit	$15.50	$15.86	$0.36
Test period, catalog and Internet profits	$19.46	$20.54	$1.08
Test and posttest periods, catalog and Internet profits	$35.06	$37.49	$2.43
Sample size	4,790	4,758	

Notes: Profits earned from each customer are calculated as the sum of the items ordered by each customer, multiplied by the profit margin on each item, minus the cost of printing and mailing catalogs during the test period.

increase in posttest demand. Customers who live a long way from a store are unlikely to make a store purchase and so have little opportunity for such channel substitution. As a result, focusing on these customers may yield a more accurate measure of the change in *overall* demand.

An alternative explanation is that customers living close and far from the stores are systematically different and these differences interacted with the outcome of the test. To investigate this possibility, we compared the historical RFM measures for both the Best and the Other customers. These measures are not significantly different for customers who live within an hour of a store and those who live more than an hour from a store. While this is reassuring, it does not rule out the possibility that there are other sources of heterogeneity among the customers, which remain unobserved. This limitation highlights the importance of the randomized experimental design. Unlike, the assignment of customers to the high and low advertising conditions, the assignment of customers to the two driving time conditions is endogenous and not random.

G. The Price of the Items Ordered

Changes in customers' purchasing behavior may be reflected both in the number of items that they order and the price of those items. Our theory does not make any predictions about how the variation in advertising frequencies will affect the price of the items that customers order. However, for completeness, we examined whether there was any difference in the average prices of the items ordered by customers in the high and low advertising conditions. We did not observe any significant differences in the average prices of the items orders, either during the test period or in the subsequent posttest period.[12]

The firm's focus is not limited to the number of items ordered or the price of those items. The firm primarily cares about profits. Our final results compare how the increase in advertising frequency affected the profits earned during the test and posttest periods.

H. Sending Catalogs to Their Best Customers

As we discussed, most companies adopt a myopic approach to their catalog mailing policies: they vary their mailing policies for a specific catalog and compare the response to that catalog. This myopic focus on short-run catalog demand ignores the externalities in other channels and in future periods. For example, the findings in Table 6 indicate that among Best customers, the short-run response to advertising overstates the long-run response to advertising by a factor of three (5.3% vs. 1.8%). Firms that base their decisions on the

12. Because the dependent measure is only defined for customers who made a purchase, we necessarily restricted this analysis to these customers. We caution that this introduces the potential for selection bias.

short-run response are likely to overinvest in advertising.

To illustrate the implications of this result on firm profit, we summarize the profits earned in each condition in Table 10. The profits are calculated as the sum of the items ordered by each customer, multiplied by the profit margin on each item, minus catalog printing and mailing costs incurred during the test period. We compare three different profit measures: (1) profit earned from the catalog channel in the test period, (2) profit earned from all channels in the test period (including Internet orders), and (3) profit earned from all channels in both the test and the posttest periods.

Focusing first on the Best customers, we see that if the company focused solely on profits earned during the test period from the catalog channel, it would erroneously conclude that it is profitable to send catalogs more frequently to its Best customers. After allowing for the adverse intertemporal and cross-channel outcomes, we see that the profit result is reversed. The company actually earned a higher average profit in the low advertising condition. Among the Other customers, the positive externalities in the Internet channel and posttest period almost lead to the opposite outcome. Mailing more frequently to the Other customers is clearly more profitable when these externalities are taken into account. However, this conclusion is much weaker if attention is restricted to test period profits from the catalog channel.

This interpretation of the findings raises the question as to why companies typically ignore these long-run and cross-channel effects. We offer two responses. First, not all catalog firms have ignored these effects. For example, Rhenania, a German book catalog company, revised its mailing policies to shift its objective function from maximizing short-run profits to also consider profits in future periods (Elsner, Krafft, and Huchzermeier 2003). The company attributed the reversal of its history of declining sales, market share, and profits to the adoption of its new mailing policy.

Our second response is that measuring and responding to long-run and cross-channel effects are difficult. Consider first the measurement problem. When customers call to place an order over the telephone, they are asked for a code printed on the catalog that identifies which catalog customers are ordering from. Similarly, when a customer orders via mail using the form bound into a catalog, companies can again identify the catalog from a code preprinted on the order form. As a result, companies can construct a rich database identifying which of the customers who received a catalog placed an order through the catalog channel. In contrast, when a customer places an order through a company's Internet Web site, it is generally not possible to identify whether the order was prompted by a catalog and (if so) which catalog the customer is ordering from. Linking purchases from future catalogs to past mailing decisions is even more difficult.

Furthermore, when future purchases are linked to past mailings as part of a controlled experiment, it is important to recognize the role of customer heterogeneity. If the Best and Other customers had been pooled in this study, then the net effect of additional advertising on future sales is statistically indistinguishable from zero. This is not because the effect on individual consumers is zero. Instead, it reflects the negative effects on the Best customers canceling out the continuing positive contributions for the Other customers. This is even more likely to be overlooked when analyzing historical data in the absence of a controlled experiment.

Even when companies can effectively measure cross-channel and long-run customer response functions, optimizing the company's mailing strategy remains difficult. Optimizing the short-run policy is relatively straightforward as there are only two possible actions: *mail* or *don't mail*. In contrast, the long-run mailing policy has an infinite range of possible mailing *sequences*. Moreover, evaluating the profitability of these sequences is no longer a straightforward statistical problem. Some catalog companies have tested sequences of mailing policies using split-sample field tests. Yet, such approaches cannot reveal the optimal policy without an infinite series of such tests.

VI. CONCLUSIONS

We have reported the findings from a large-scale field study in which we exogenously manipulated the frequency of catalog advertising sent to randomly selected customer samples. We then tracked both the immediate response and the impact on future purchases by these customers. The findings confirm that retail advertising can impact future demand, but surprisingly, the sign of the impact varies across customers. Among the company's most valuable customers, who had purchased recently and frequently from the company, the long-run impact was negative. The short-run lift in demand for these customers was apparently largely due to cross-channel and temporal substitutions.

Among the less valuable customers, who had purchased less frequently and/or less recently, there is evidence that advertising had a positive impact on future demand. However, this outcome was limited to customers living close to one of the firm's retail stores, suggesting that the result may provide further evidence of channel substitution.

The findings offer an explanation for a question that has often left customers perplexed: why do companies send so many catalogs to their best customers? It seems that the intensive mailing frequency to a company's best customers can be explained in part by a (mistaken) focus on short-run outcomes when designing catalog mailing policies. If a company overlooks the negative externalities on future demand and demand in other channels, it will tend to overmail to its best customers. The same myopic focus may lead to the opposite outcome for other "less valuable" customers. For these customers, the externalities are positive, so that it may be profitable to mail to customers who are unlikely to purchase immediately, as by doing so, companies can increase the probability of a future purchase.

We conclude that advertising can cause both increases and decreases in future demand, and the outcome depends on the characteristics of the customers. Our results also demonstrate the power of field experiments not only for advancing research on the economics of advertising but also in identifying potential gaps in management practice.

APPENDIX

We consider the following example to illustrate the long-run effects of advertising. Assume that utility is a separable quadratic function:

$$U(q_1, q_2, \bar{q}) = q_1(v_1 - q_1) + q_2(v_2 - q_2) + \bar{q}(\bar{v} - \bar{q}). \tag{2}$$

To simplify exposition, we normalize all prices to one, which is analogous to assuming a physical constraint. For example, if customers have a limit on the size of their wardrobe, there may be a physical constraint on how many new clothes they can purchase during the course of a season. We assume that customers will always prefer to choose q_t less than v_t and we make analogous assumptions for the competitive product.

Solving the resulting system of first-order conditions reveals customers' optimal consumption decisions:

$$q_1^* = (2v_1 - v_2 + 2Y - \bar{v})/6 \tag{3}$$

$$q_2^* = (2v_2 - v_1 + 2Y - \bar{v})/6 \tag{4}$$

$$\bar{q} = (2\bar{v} - v_1 - v_2 + 2Y)/6. \tag{5}$$

The key insights concern the relationship between advertising in Period 1 and customers' purchasing decisions of the focal company's products.

$$\mathrm{d}q_1^*/\mathrm{d}a_1 = (2\mathrm{d}v_1/\mathrm{d}a_1 - \mathrm{d}v_2/\mathrm{d}a_1)/6 > 0 \tag{6}$$

$$\mathrm{d}q_2^*/\mathrm{d}a_1 = (2dv_2/\mathrm{d}a_1 - \mathrm{d}v_1/\mathrm{d}a_1)/6. \tag{7}$$

As we would expect, the impact of Period 1 advertising on Period 1 demand is positive: $\mathrm{d}q_1^*/\mathrm{d}a_1 > 0$. The impact on future demand (q_2^*) is ambiguous.

Now, consider a segment of customers whose preferences for the focal firm are so strong that they do not purchase any units from the competing firm. After setting $\bar{q}^* = 0$ and maximizing utility subject to $Y = q_1 + q_2$, the first-order condition for q_2 yields the following second period demand:

$$q_2^* = (v_2 - v_1 + 2Y)/4. \tag{8}$$

Among consumers who never purchase the outside goods, the long-run impact of advertising is no longer ambiguous: $\mathrm{d}q_2^*/\mathrm{d}a_1 \leq 0$. Sending additional advertising to these customers cannot lead to any further interfirm substitution, and so the only remaining effect is intertemporal substitution. In contrast, among customers with weaker exante preferences for the firm, if the carryover effect of advertising is large ($2\mathrm{d}v_2/\mathrm{d}a_1 > \mathrm{d}v_1/\mathrm{d}a_1$), then the long-run effect of advertising is positive.

TABLE A1
Comparison of Test Period, Posttest Period, and Overall Results

	Posttest Period		Overall Period	
	Other Customers	**Best Customers**	**Other Customers**	**Best Customers**
High advertising	0.100** (0.026)	−0.036** (0.013)	0.125** (0.016)	0.018* (0.008)
Recency	−0.288** (0.008)	−0.146** (0.005)	−0.281** (0.005)	−0.138** (0.003)
Frequency	0.465** (0.014)	0.723** (0.010)	0.478** (0.008)	0.737** (0.006)
Monetary value	0.187** (0.037)	0.515** (0.032)	0.335** (0.023)	0.713** (0.020)
Driving time	−0.011 (0.012)	0.002 (0.006)	−0.013 (0.007)	0.004 (0.004)
Intercept	−0.746** (0.166)	−3.195** (0.146)	−0.367 (0.105)	−3.195** (0.092)
Sample size	9,458	9,761	9,458	9,761

Notes: The posttest findings report the coefficients from Equation (1) estimated using data from the posttest period. The overall period findings report the coefficients from Equation (1) estimated using data from the entire period (test and posttest). Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

TABLE A2
Comparison of Posttest Results, Start and End of the Posttest Period

	Start of Posttest Period		End of Posttest Period	
	Other Customers	**Best Customers**	**Other Customers**	**Best Customers**
High advertising	0.114** (0.038)	−0.096** (0.019)	0.087* (0.037)	0.021 (0.018)
Recency	−0.273** (0.012)	−0.162** (0.008)	−0.302** (0.011)	−0.131** (0.007)
Frequency	0.447** (0.020)	0.755** (0.014)	0.485** (0.019)	0.692** (0.013)
Monetary value	0.298** (0.055)	0.691** (0.047)	0.091 (0.048)	0.341** (0.045)
Driving time	0.013 (0.017)	0.015 (0.008)	−0.034* (0.017)	−0.011 (0.008)
Intercept	−1.955** (0.252)	−4.656** (0.209)	−1.000** (0.217)	−3.124** (0.202)
Sample size	9,458	9,834	9,458	9,834

Notes: The findings report the coefficients from Equation (1) estimated using purchases made at the start and end of the posttest period. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

TABLE A3
Comparison of Test Period Results by Channel

	Internet Channel		Catalog Channel	
	Other Customers	**Best Customers**	**Other Customers**	**Best Customers**
High advertising	0.281** (0.055)	−0.092* (0.038)	0.118** (0.020)	0.065** (0.011)
Recency	−0.454** (0.016)	−0.066** (0.016)	−0.247** (0.007)	−0.138** (0.004)
Frequency	0.567** (0.028)	0.829** (0.028)	0.474** (0.011)	0.741** (0.008)
Monetary value	0.270** (0.073)	1.354** (0.094)	0.444** (0.031)	0.801** (0.027)
Driving time	−0.076** (0.025)	−0.088** (0.017)	−0.006 (0.009)	0.014** (0.005)
Intercept	−2.113** (0.330)	−9.527** (0.425)	−1.526** (0.145)	−4.173** (0.122)
Sample size	9,458	9,761	9,458	9,761

Notes: The findings report the coefficients from Equation (1) estimated using test period purchases through each channel. Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

TABLE A4

Comparison of Test Period, Posttest Period, and Overall Results: Customers Living More Than an Hour from a Store

	Test Period		Posttest Period		Overall Period	
	Other Customers	**Best Customers**	**Other Customers**	**Best Customers**	**Other Customers**	**Best Customers**
High advertising	0.163** (0.023)	0.082** (0.013)	0.020 (0.032)	−0.048** (0.016)	0.113** (0.019)	0.030** (0.010)
Recency	−0.286** (0.007)	−0.133** (0.005)	−0.273** (0.010)	−0.151** (0.006)	−0.282** (0.006)	−0.140** (0.004)
Frequency	0.457** (0.012)	0.738** (0.009)	0.489** (0.016)	0.708** (0.011)	0.468** (0.010)	0.726** (0.007)
Monetary value	0.456** (0.035)	0.915** (0.031)	0.210** (0.045)	0.535** (0.039)	0.369** (0.028)	0.766** (0.024)
Driving time	0.002 (0.019)	0.033** (0.010)	0.061* (0.026)	−0.060** (0.013)	0.022 (0.016)	−0.004 (0.008)
Intercept	−1.249** (0.161)	−4.601** (0.139)	−0.981** (0.208)	−3.141** (0.173)	−0.500** (0.128)	−3.355** (0.108)
Sample size	6,555	6,628	6,555	6,628	6,555	6,628

Notes: The posttest findings report the coefficients from Equation (1) estimated using data from the posttest period. The overall period findings report the coefficients from Equation (1) estimated using data from the entire period (test and posttest). Standard errors are given in parentheses.

*Significantly different from zero, $p < .05$; **significantly different from zero, $p < .01$.

REFERENCES

Aaker, D., and J. M. Carman. “Are Your Overadvertising?” *Journal of Advertising*, 22(4), 1982, 57–70.

Ackerberg, D. A. “Empirically Distinguishing Informative and Prestige Effects of Advertising.” *Rand Journal of Economics*, 32, 2001, 316–33.

———. “Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination.” *International Economic Review*, 44, 2003, 1007–40.

Anderson, E. T., and D. I. Simester. “Impact of Promotion Depth on New vs. Established Customers.” *Marketing Science*, 23(1), 2004, 4–20.

Bagwell, K. Forthcoming. “The Economic Analysis of Advertising, ” in *Handbook of Industrial Organization*, Vol. 3, edited by M. Armstrong and R. Porter. Amsterdam, The Netherlands: North-Holland, 2005, 1701–1844.

Becker, G. S., and K. M. Murphy. “A Simple Theory of Advertising as a Good or Bad.” *Quarterly Journal of Economics*, 108, 1993, 942–64.

Benham, L. “The Effects of Advertising on the Price of Eyeglasses.” *Journal of Law and Economics*, 15, 1972, 337–52.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman. “What’s Psychology Worth? A Field Experiment in the Consumer Credit Market.” Working Paper, University of Chicago, 2006.

Blattberg, R., and S. Neslin. *Sales Promotions*. Englewood Cliffs, NJ: Prentice Hall, 1990.

Borden, N. H. *The Economic Effects of Advertising*. Chicago: Richard D. Irwin, 1942.

Comanor, W. S., and T. A. Wilson. “Advertising, Market Structure and Performance.” *Review of Economics and Statistics*, 49, 1967, 423–40.

———. *Advertising and Market Power*. Cambridge, MA: Harvard University Press, 1974.

Deighton, J., C. M. Henderson, and S. A. Neslin. “The Effects of Advertising on Brand Switching and Repeat Purchasing.” *Journal of Marketing Research*, 31(1), 1994, 28–43.

Direct Marketing Association. *The DMA 2003 Statistical Fact Book*. New York: Direct Marketing Association, 2003.

Elsner, R., M. Krafft, and A. Huchzermeier. “Optimizing Rhenania’s Mail-Order Business through Dynamic Multilevel Modeling.” *Interfaces*, 33(1), 2003, 50–66.

Erdem, T., and M. Keane. “Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets.” *Marketing Science*, 15(1), 1996, 1–20.

Grossman, G., and C. Shapiro. “Informative Advertising with Differentiated Products.” *Review of Economic Studies*, 51, 1984, 63–81.

Harrison, G. W., and J. A. List. “Field Experiments.” *Journal of Economic Literature*, 42, 2004, 1009–55.

Hendel, I., and A. Nevo. “Sales and Consumer Inventory.” National Bureau of Economic Research Working Paper No. 9048. Cambridge, MA: National Bureau of Economic Research, 2002.

———. “The Post Promotion Dip Puzzle: What do the Data Have to Say?” *Quantitative Marketing and Economics*, 1, 2003, 409–24.

———. “Measuring the Implications of Sales and Consumer Inventory Behavior.” National Bureau of Economic Research Working Paper No. 11307. Cambridge, MA: National Bureau of Economic Research, 2005.

Ippolito, P. M., and A. D. Mathios. “Information, Advertising and Health Choices: A Study of the Cereal Market.” *RAND Journal of Economics*, 21, 1990, 459–80.

Kaldor, N. “The Economic Aspect of Advertising.” *Review of Economic Studies*, 18, 1950, 1–27.

Kihlstrom, R., and M. Riordan. “Advertising as a Signal.” *Journal of Political Economy*, 92, 1984, 427–50.

Krishnamurthi, L., and S. P. Raj. “The Effect of Advertising on Consumer Price Sensitivity.” *Journal of Marketing Research*, 22, 1985, 119–29.

Lambin, J. J. *Advertising, Competition and Market Conduct in Oligopoly Over Time*. Amsterdam, The Netherlands: North-Holland, 1976.

Lodish, L. M., M. M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, and

M. E. Stevens. "How T. V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments." *Journal of Marketing Research*, 32(2), 1995a, 125–39.

Lodish, L. M., M. M. Abraham, J. Livelsberger, B. Lubetkin, B. Richardson, and M. E. Stevens. "A Summary of Fifty-Five In-Market Experimental Estimates of the Long-Term Effect of TV Advertising." *Marketing Science*, 14(3), Part 2, 1995b, G133–40.

Milgrom, P., and J. Roberts. "Price and Advertising Signals of Product Quality." *Journal of Political Economy*, 94, 1986, 796–821.

Milyo, J., and J. Waldfogel. "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart." *American Economic Review*, 89(5), 1999, 1081–96.

Nelson, P. "Information and Consumer Behavior." *Journal of Political Economy*, 78, 1970, 311–29.

———. "Advertising as Information." *Journal of Political Economy*, 82, 1974, 729–54.

Roberts, M. J., and L. Samuelson. "An Empirical Analysis of Dynamic Non-Price Competition in an Oligopolistic Industry." *Rand Journal of Economics*, 19, 1988, 200–20.

Schmalensee, R. *"The Economics of Advertising."* Amsterdam, The Netherlands: North-Holland, 1972.

———. "A Model of Advertising and Product Quality." *Journal of Political Economy*, 86, 1978, 485–503.

Seldon, B. J., and K. Doroodian. "A Simultaneous Model of Cigarette Advertising: Effects on Demand and Industry Response to Public Policy." *Review of Economics and Statistics*, 71, 1989, 673–77.

Stigler, G. J. "The Economics of Information." *Journal of Political Economy*, 69, 1961, 213–25.

Tellis, G. J. "Advertising Exposure, Loyalty and Brand Purchase: A Two-stage Model of Choice." *Journal of Marketing Research*, 25(2), 1988, 134–44.

Telser, L. G. "Advertising and Competition." *Journal of Political Economy*, 72, 1964, 537–62.

Wells, W. D., S. Moriarity, and J. Burnett. *Advertising: Principles and Practice*. Englewood Cliffs, NJ: Prentice Hall, 2006.