Does Current Advertising Cause Future Sales?
Evidence from the Direct Mail Industry

December, 2005

Duncan Simester
MIT Sloan School of Management
simester@mit.edu

Yu (Jeffrey) Hu
Purdue University
yuhu@purdue.edu

Erik Brynjolfsson
MIT Sloan School of Management
erikb@mit.edu

Eric T. Anderson
Northwestern University
eric-anderson@kellogg.northwestern.edu

Abstract

We present findings from a large-scale field experiment that allows us to study whether there is a causal relationship between current advertising and future sales. The experimental design overcomes limitations that have affected previous investigations of this issue. We find that current advertising does affect future sales but the sign of the effect varies depending on the customers targeted. For the firm's best customers the long-run effect of increases in current advertising is actually negative, while for other customers the effect is positive. We argue that these outcomes reflect two competing effects: brand-switching and inter-temporal substitution. Furthermore, our data suggest a way to distinguish between the informative and persuasive roles of advertising, providing insight into the mechanism by which advertising differentially affects various customer subsets.

Acknowledgements: 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, Fifth Annual INFORMS Revenue Management and Pricing Conference. Generous funding was provided by MIT Center for eBusiness.

1. Introduction

Over $245 billion was spent on advertising in the United States in 2003 (Advertising Age, 2005). Despite enormous levels of spending, an important economic question has yet to be resolved. Is there a positive association between current advertising and future sales? Years of academic research investigating the relationship between advertising and future demand has yielded inconclusive findings. Most of these previous studies have been plagued by two obstacles:

1. Advertising decisions are endogenous and so effects attributed to variations in advertising expenditure may actually reflect factors that led to the variation in expenditure.
2. Advertising is dynamic and so the effects of advertising are often confounded by other intervening effects.

These obstacles are well-recognized and a variety of econometric methods have been proposed to address them. In this paper we report findings from a large–scale field test in which we address both problems using an alternative approach. We experimentally vary advertising strategies for two randomly selected Treatment and Control samples of customers. This experimental manipulation introduces an exogenous source of variation that overcomes the endogeneity limitation that has limited previous studies. Both samples are exposed to the same intervening events such as competitive reactions or macro-economic changes. Thus, the comparison between the Treatment and Control samples overcomes these potential confounds.

The field test was conducted in the direct mail industry with a mail-order catalog that sells women's clothing in the moderate price range. Direct mail represents 20% of all advertising in the United States and is the single largest type of media advertising (Advertising Age, 2005). As a comparison, 2003 advertising spending was $48 billion in direct mail, $45 billion in newspapers and $42 billion in broadcast television. In the United States direct mail companies mailed over 18 billion catalogs in 2004 (Direct Marketing Association, 2005). The catalogs typically announce a company's retail prices and contain detailed information about the available products, together with information about ordering procedures, warranties, and payment methods. This focus on providing information about available products and prices is typical of advertising by retailers (and contrasts with advertising by manufacturers).

It is well accepted in the direct mail industry that catalog advertising increases current demand. What is much less well-understood is the impact that it has on demand in *future* periods. With few exceptions, companies adopt a myopic focus when optimizing their catalog advertising decisions. They estimate the probability that a customer will order from a specific catalog and then mail to all customers for whom the expected lift in immediate demand justifies the incremental printing and mailing costs. This focus solely on the current impact of catalog advertising either implicitly presumes that there is no long-term effect or simply ignores any long-term effects.

The possibility that catalog advertising may have a long-run impact on demand has been recognized in the academic literature. Indeed, models have been proposed to help companies solve the difficult dynamic optimization problem that arises if catalogs do impact future demand (see for example Gönül and Shi, 1998; and Simester, Sun and Tsitsiklis 2005). However, there are apparently no published studies directly estimating the impact of catalog advertising on future demand or comparing how these long-run effects (if any) vary across customers.

1.1 Prior Theoretical Work

Does current advertising increase future sales? Unfortunately, existing theory and evidence provides two conflicting answers to this basic question. Much of the theoretical advertising literature has focused on distinguishing whether advertising serves a persuasive or informative role. Under the "persuasive" view, advertising alters customers' utility functions by changing their preferences (Kaldor 1950; Comanor and Wilson 1967 and 1974; and Becker and Murphy 1993). This leads to an outward shift in the demand function, which has led to claims that advertising may serve an important anti-competitive role. Under the "informative" view, advertising does not change customers' utility functions (Stigler 1961; Kihlstrom and Riordan 1984; and Milgrom and

Roberts 1986).[1] Instead it increases the information that customers have about the available alternatives.

Both views of the role of advertising predict an increase in customers' expected utility from consuming the advertised product. Under the persuasive view the change in expected utility reflects a change in the utility function itself, while the informative view predicts a change in which alternatives are evaluated and/or what is known about those alternatives. Both outcomes are consistent with advertising positively impacting demand in future periods. In particular, under the persuasive view we would generally expect a change in the utility function to endure into future periods. Similarly, product information revealed under the informative view will generally be relevant in future periods, unless changes between periods make past information obsolete.

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 outcomes, the impact on competing brands (sometimes termed the "combative" role of advertising) has received the most interest. As early as 1942 Borden distinguished between the "primary" and "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 tobacco industry (see for example Seldon and Doroodian 1989; and Roberts and Samuelson 1988). 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, anti-

[1] See also: Telser 1964; Nelson 1970 and 1974; Schmalensee 1978; and Grossman and Shapiro 1984.

smoking 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 inter-temporal 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 inter-temporal substitution. As a result, following a price promotion, there is often evidence of a "post-promotion dip" in sales, as customers consume products purchased during the discount period (Blattberg and Neslin 1990, p. 358; and Hendel and Nevo 2003).[2] Interestingly, there is also evidence that this inter-temporal 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 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, temporal-substitution may lead to a negative impact on future demand.

1.2 Prior Empirical Evidence

The previous empirical evidence is mixed. There is some evidence of a positive long-run relationship between advertising and sales. Yet many studies report either no long-run impact or that the impact is short-lived (Bagwell 2005). There have apparently not been any studies reporting a negative relationship between advertising and future demand. However, as we recognized, this empirical work has been confronted by important challenges. The earlier work was typically limited to aggregate brand or category-level

[2] See also Hendel and Nevo (2002 and 2005).

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 this relationship even when it is present. These studies also suffered from important limitations due to the endogeneity of the advertising decisions, since changes in sales can lead to changes in advertising budgets, and confounds introduced by intervening events (Schmalensee 1972; and Lambin 1976). More recently, the development of household level panel datasets has made it possible to estimate demand at the individual or household-level. Together with methodological developments in the estimation of simultaneous structural models, these new datasets offer the opportunity to address endogeneity through advanced econometric controls (see for example: Erdem and Keane 1996; and Ackerberg 2003).

In contrast, our approach has not been to exploit ever more sophisticated econometric methods but rather to improve the direct measures of advertising and sales. In particular, the direct sales industry provides a particularly measurable domain for studying responses to advertising. Furthermore, the experimental approach that we adopt in this paper departs from most earlier attempts to overcome endogeneity and intervening effects by introducing random assignment of customers to “Treatment” and “Control” groups with external controls to the data collection process to 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 treatment. We directly measure the difference in their long-run demand. The outcome is immediately interpretable.

This is not the first field experiment designed to investigate the impact of advertising. Managerial studies using proprietary split-sample cable TV 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 and estimation models (see for example: Aaker and Carmen 1982; Lodish et al. 1995a and 1995b). Moreover, the results are apparently mixed, perhaps in part due to a lack of statistical power.[3] In addition, there have been at least two academic studies that use experiments to investigate how advertising influences prices and price elasticities. Krishnamurthi and Raj (1985) also report the findings from a split sample cable TV 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 advertising on prices. They find that advertising does tend to lower the prices of advertised products, but has little effect on the prices of unadvertised products.[4]

1.3 Preview of the Key Findings

The field experiment reported in our paper reveals several findings. First, as we would expect, customers in the Treatment condition who received additional catalogs purchased more items in the short-term than their counterparts in the Control condition. This result confirms that current advertising can increase current demand. As we discussed, this finding is already well-understood in the direct mail industry but is also the limit of most firms' analyses. These firms do not look beyond the immediate impact to also consider how catalog advertising affects demand in the future or demand in other channels.

Our findings confirm that current advertising also has a significant impact on demand in future periods. However, the effect is not always positive. Among the catalogs' "Best" customers, who had historically purchased recently and frequently, increased current advertising significantly *reduced* future demand. We interpret this result as evidence of temporal substitution. In fact, we find that the short-run increase in demand among these customers is almost entirely offset by the reduction in future demand. Further investigation also revealed evidence of cross-channel substitution: the increase in demand

---

[3]These three papers do not report sample sizes or the estimation models for individual studies.
[4] For other examples of natural experiments see: Benham (1972) and Ippolito and Mathios (1990).

from the catalog channel (mail and telephone orders) was offset by a reduction in demand from the Internet channel.

In contrast, for the "Other" customers in the study, who had historically purchased less frequently and/or less recently than the Best customers, the results were reversed. Sending more current catalogs led to an increase in future demand.

Why do companies mail so many catalogs to their best customers? Our findings may provide an explanation. It is not uncommon for companies to send catalogs every two weeks, with some companies sending their best customers as many as 100 catalogs a year. These intensive mailing policies often prompt complaints from customers that they receive too many catalogs. Yet paper and postage are not free and so these policies are not arbitrary. One explanation is that firms are myopic and limit their attention to the short-run impact of mailing decisions on purchases from the catalog channel. If companies were to extend their analysis to also consider the long-run and cross-channel impacts, they would learn that for their best customers, the lift in short-run demand is largely due to inter-temporal and cross-channel substitution. In later discussion we offer an explanation for why companies commonly overlook these externalities.

1.4 Structure of the Paper

The paper proceeds in Section 2 with a simple model illustrating the intuition that current advertising may lead to a positive or a negative impact on future demand. We then provide an overview of the study design in Section 3 before presenting the results in Section 4. 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 paper concludes in Section 5 with a review of the findings and implications.

2. Positive and Negative Long-Run Outcomes

To help understand why current advertising may lead to a positive or a negative impact on future demand we present a stylized model that highlights two opposing advertising outcomes: brand-switching and inter-temporal substitution.

We consider a two period problem in which a firm produces different products each period. In the first period customers decide how many products to purchase and consume in that period (q_1) and how many products to purchase and stockpile for the next period (i_1). In the second period, customers choose how many additional units to purchase and consume in that period (q_2). Stockpiling between periods allows customers to introduce variety to their second period consumption decisions (recall that the firm produces different products each period). For simplicity we assume that the price charged by the focal firm (p) does not vary between periods and set the inter-period discount rate to 1. These assumptions jointly ensure that in this model stockpiling cannot be explained by mere price arbitrage.

Customers can also choose to purchase a competing brand. We use an upperbar to distinguish the price of the competing brand ($\overline{p}$) and the customers' quantity decisions ($\overline{q}_t$) for this competitor. We set the competitor's price to one and do not consider stock-piling of this outside option as it adds few additional insights and does not alter our key findings. Finally, we also assume that there is a budget constraint such that: $Y = p\left(q_1 + i_1 + q_2\right) + \overline{p}\left(\overline{q}_1 + \overline{q}_2\right)$.

To simplify the analysis and exposition it is helpful to describe consumption utility using a separable quadratic function:

$$U\left(q_1, \overline{q}_1\right) = q_1\left(v_1 - q_1\right) + \overline{q}_1\left(\overline{v} - \overline{q}_1\right) \tag{2}$$

$$U\left(i_1, q_2, \overline{q}_2\right) = i_1\left(v_1 - i_1\right) + q_2\left(v_2 - q_2\right) + \overline{q}_2\left(\overline{v} - \overline{q}_2\right) \tag{3}$$

The v_t terms are preference parameters that are influenced by advertising.[5] 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: $dv_t / da_t > dv_t / da_{t-j} > 0 \;\forall\; j > 0$, while $dv_t / da_{t+k} = 0 \;\forall\; k > 0$. We also assume that advertising by the focal firm does not directly affect preferences for the competing product: $d\overline{v} / da_t = 0$. While the relationship between advertising and preferences for the focal brand is positive, this illustrative model does not speak to the source of this relationship. In particular, we do not seek to distinguish between the *information* and *persuasion* interpretations proposed in the literature. We later use our empirical findings to investigate this issue.

Customers select the quantity of goods that maximizes utility for both periods, subject to their budget constraint. Solving the resulting system of first-order conditions reveals customers' optimal consumption decisions:

$$q_1^* = i_1^* = \frac{3v_1 - v_2 + 2(Y - \overline{v})}{10} \tag{4}$$

$$q_2^* = \frac{2v_2 - v_1 + (Y - \overline{v})}{5} \tag{5}$$

$$\overline{q}_1^* = \overline{q}_2^* = \frac{3\overline{v} + 2Y - (2v_1 + v_2)}{10} \tag{6}$$

The key insights concern the relationship between advertising in period 1 and customers' purchasing decisions of the inside goods.

$$\frac{dq_1^*}{da_1} = \frac{di_1^*}{da_1} = \frac{1}{3}\left(3\frac{dv_1}{da_1} - \frac{dv_2}{da_1}\right) > 0 \tag{7}$$

$$\frac{dq_2^*}{da_1} = \frac{1}{3}\left(2\frac{dv_2}{da_1} - \frac{dv_1}{da_1}\right) \tag{8}$$

As we would expect, the impact of period 1 advertising on period 1 demand is positive: $dq_1^* / da_1 > 0$ and $di_1^* / da_1 > 0$. The impact on future demand (q_2^*) is ambiguous and

---

[5] We assume that $2Y \le v_t \;\forall t$, which ensures that $2q_t \le v_t \;\forall t$ and therefore consumption utility is always increasing in quantity. In practice, as long as $p > 0$ customers will always prefer to choose q_t and i_t less than v_t. We also make analogous assumptions for the competitive product.

reflects a trade-off between brand-switching and inter-temporal substitution. In particular, the sign of $dq_2^* / da_1 > 0$ depends upon the rate at which the long-run impact of advertising decays. Because advertising in period 1 has a favorable impact on preferences in period 2 ($dv_2 / da_1 > 0$) it leads to a switch in second period demand from the competing brand to the focal brand. Yet current advertising has a bigger impact on current preferences than future preferences ($dv_2 / da_1 < dv_1 / da_1$), and so the lift in second period demand is offset by forward-buying in which customers shift second period demand to the first period.

This analysis also suggests that the relationship between current advertising and future demand will vary across customer segments. If customers already have strong preferences for the brand, there is relatively little opportunity for additional brand-switching. For example, consider a segment of customers whose preferences for the focal firm are so strong that they do not purchase any units from the competing brand. After setting $\overline{q}_1^* = \overline{q}_2^* = 0$ and maximizing utility subject to $Y = \left(q_1 + i_1 + q_2\right)$, the first order condition for q_2 yields the following second period demand:

$$q_2^* = \frac{v_2 - v_1 + Y}{3} \tag{9}$$

Among consumers who never purchase the outside goods, the long-run impact of advertising is no longer ambiguous: $dq_2^* / da_1 \le 0$. Sending additional advertising to these customers cannot lead to any further brand-switching, and so the only remaining effect is inter-temporal substitution. We conclude that for customers with very strong preferences for the firm, current advertising may lead to a reduction in future demand. In contrast, among customers with weaker *ex ante* preferences for the firm, the possibility of brand switching is more likely to lead to a favorable long-run outcome.

The study described in the next section provides an opportunity to test these predictions. We can use customers' transaction histories (prior to the study) to distinguish customers with strong preferences for the focal firm from those with weaker preferences. Random assignment yields equivalent Treatment and Control samples of both types of customers.

Therefore, comparing how the response to the experimental manipulation varied across these samples reveals how prior preferences moderate the long-run impact of advertising.

3. Study Design

The study was conducted with a medium-sized company that sells women's clothing in the moderate price range.[6] All of the products carry the company's private label brand and are sold exclusively through the company's own catalogs, Internet website and retail stores. The study involved a total of 20,000 customers who had previously made a mail or telephone purchase from the company. 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.[7] The "Other" sample of 10,000 customers was comprised of customers who the company's statistical model predicted had an average probability of responding if mailed a catalog.

Random assignment was then used to assign these two samples of customers into equal sized Treatment and Control groups. This yielded a total of four different customer samples (see 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.[8]

---

[6] The company asked to remain anonymous.
[7] Although the details of the company's statistical models are proprietary and were not made available to the research team, the recency and frequency of prior purchases accurately distinguish these customers.
[8] 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

	Control Sample	**Treatment Sample**
“Best” Customers	4,921	4,904
“Other” Customers	4,790	4,758

The experimental manipulation occurred over an (approximately) eight-month period. During this period all of the customers in the Treatment sample received a total of seventeen catalogs, while customers in the Control sample received just twelve catalogs. The additional catalogs sent to the Treatment 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. The cost of designing new catalogs is expensive and so rather than designing new catalogs companies will often re-send the same catalog two to four weeks after the first mailing.

The actual mailing schedule for the two samples is summarized in Table 2. The specific timing of each mailing was determined by the company’s circulation managers. The managers were instructed to optimize the overall (short-run) response given the exogenous decision to mail a total of twelve times to the Control sample and seventeen times to the Treatment samples. It is possible that varying the timings would lead to differences in the long-run results. Following the experimental manipulations the company returned to using its standard circulation procedures to decide who to mail catalogs to, and made no distinction between customers in the Treatment and Control samples.

Table 2: Mailing Dates in 2002 by Experimental Condition

	Control	Treatment
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

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 Treatment group were sent Catalog 2). The last date on which the mailing dates 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 nineteen months 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 days beyond the date of the last manipulation: September 20, 2002 through December 31, 2002. This was designed to capture orders from catalogs mailed towards the end of the manipulation period. The company estimated that over 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 DMA (Direct Marketing Association 2003). We later vary the length of the Posttest period to investigate how it affects the results (see Table 6).

We caution that the transaction data only involves customers’ purchases through the company’s Internet website or its catalog channel (mail and telephone orders). 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 discuss how this omission may have affected the results.

The historical purchasing results provide a means of checking whether the assignment of customers to the Treatment and Control conditions was truly random. In particular, in Table 3 we compare the average *Recency*, *Frequency* and *Monetary Value* (RFM) of customers’ purchases during the Pretest period.[9] If the random assignment was truly random we should not observe any systematic differences in these historical measures between the Treatment and Control samples. The findings reveal no significant differences in the historical demand in either the Best-customer or Other-customer comparisons.

[9] “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 3: Check on Randomization Process
Purchases During the Pretest Period

	Control Condition	Treatment Condition	p-value
Best Customers			
Recency	1.43 (0.02)	1.43 (0.01)	0.72
Frequency	40.38 (0.45)	40.75 (0.51)	0.59
Monetary Value	61.11 (0.19)	61.22 (0.19)	0.69
Sample Size	4,921	4,904	
Other Customers			
Recency	4.67 (0.06)	4.76 (0.06)	0.30
Frequency	10.56 (0.20)	10.62 (0.21)	0.85
Monetary Value	63.85 (0.29)	64.18 (0.33)	0.50
Sample Size	4,790	4,758	

The table reports the average values of each variable for each sub-sample. Standard errors are in parentheses. The p-value denotes the probability that the difference between the Treatment and Control averages will be larger than the observed difference (under the null hypothesis that the true averages are identical).

4. Results

4.1 Does Current Advertising Impact Short-Run Demand?

In Table 4 we summarize demand in the Treatment and Control conditions during the Test period and report both univariate and multivariate comparisons. The univariate analysis is simply the average number of items purchased by customers in each sample. The multivariate analysis uses customers' pretest purchases to control for individual customer characteristics. In particular, the Recency, Frequency and Monetary Value (RFM) of customers' prior purchases, which we used to check the validity of the randomization procedures (see Table 3), are well-established metrics for segmenting customers in this industry and provide natural candidates for control variables. The unit of analysis in the multivariate analysis 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. In particular, we assume that Q_i is drawn from a Poisson distribution with parameter λ_i:

$$\text{Prob}\left(Q_i = q\right) = \frac{e^{-\lambda_i}\lambda_i^q}{q!}, \qquad q=0, 1, 2, \ldots \tag{10}$$

where: $\ln\left(\lambda_i\right) = \boldsymbol{\beta}\mathbf{X}_i$. The $\mathbf{X}_i$ terms denote the independent variables, which include the log of each of the three RFM measures. To evaluate the impact of the experimental manipulation we include a dummy variable identifying whether customer i was in the Treatment condition. This yields the following model:

$$\begin{aligned} \boldsymbol{\beta}\mathbf{X}_i = {} & \beta_0 + \beta_1 \log(\mathit{Recency}_i) + \beta_2 \log(\mathit{Frequency}_i) + \beta_3 \log(\mathit{Monetary\ Value}_i) \\ & + \beta_4 \mathit{Treatment}_i \end{aligned} \tag{11}$$

This analysis preserves the benefits of the experimental design. Under this specification, β_4 measures the percentage change in short-run demand between customers in the Treatment condition compared to those in the Control. This comparison with the Control provides an explicit control for intervening factors, such as competitors' actions and macro-economic factors. We estimated separate models for the *Best* and *Other* customers.

Table 4: Units Ordered During the Test Period

	Other Customers	Best Customers
Univariate Analysis		
Control Condition	1.08 (0.04)	3.63 (0.08)
Treatment 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		
Intercept	-1.213** (0.133)	-4.255** (0.117)
Recency	-0.276** (0.006)	-0.131** (0.004)
Frequency	0.489** (0.010)	0.749** (0.008)
Monetary Value	0.424** (0.029)	0.827** (0.026)
Treatment	0.138** (0.019)	0.051** (0.010)
Log Likelihood	-19,160	-33,919
Sample Size	9,548	9,825

The univariate analysis reports the average number of units purchased during the Test period. The multivariate analysis reports the coefficients from Equation 11. Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$. *Significantly different from zero, $p < 0.05$.

The findings reveal that the additional advertising received by the Treatment sample led to a significant short-run increase in demand for both the *Best* and *Other* customers. The demand increase was approximately 5.1% 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 small base. In absolute terms the effect was not significantly different across the two populations. We conclude that current advertising can lead to a significant increase in short-run demand.

While these results are reassuring, they are not the main focus of this paper. Instead, we are interested in learning how increasing current advertising affects demand in future periods.

4.2 Does Current Advertising Impact Future Demand?

We report the impact of the Treatment on Posttest demand in Table 5. For the sake of brevity we restrict attention to the multivariate analysis and only report the coefficients for the *Treatment* variable. Complete findings are reported in Table A5 in the Appendix. As a basis of comparison we repeat the corresponding coefficients for the Test period and also report the coefficients when combining the data from both the Test and Posttest periods (we label this the "Total" period).

Table 5: Comparison of Test Period, Posttest Period and Total Results

	Other Customers	**Best Customers**
Test Period	0.138** (0.019)	0.051* (0.010)
Posttest Period	0.097** (0.026)	-0.037** (0.013)
Total: Test and Posttest Periods	0.124** (0.015)	0.016* (0.008)
Sample Sizes	9,548	9,825

The table reports the *Treatment* variable coefficients when estimating Equation 11 separately on the Test period, Posttest period and Total period datasets. Complete findings (including the omitted coefficients) are reported in Table A5 in the Appendix. Standard errors are in parentheses.
**Significantly different from zero, $p < 0.01$. *Significantly different from zero, $p < 0.05$.

The findings reveal a strikingly different picture for the *Best* and *Other* customers. Amongst the *Other* customers the demand expansion during the Test period persists throughout the Posttest period. The effect size drops from 13.8% in the Test period to 9.7% in the Posttest period, but remains significantly different from zero. Amongst the *Best* customers we also see a significant long-run effect, however, the sign of the effect is reversed. The increase in demand during the Test period in the Treatment condition is offset by a significant reduction in Posttest demand. This finding for the *Best* customers is consistent with temporal substitution. The increase in demand during the Test period appeared to result at least in part from substitution, with customers shifting purchases from the Posttest period to the Test period. While similar patterns of results have been reported for price promotions, to our knowledge this is the first evidence of a significant negative long-run effect attributed to advertising. We note that the findings cannot be attributed to price differences as we only manipulated the frequency with which catalogs were mailed, and not the content of the catalogs.

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-day) sub-periods 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 sub-periods are summarized in Table 6 (detailed findings are presented in Table A6 in the Appendix).

Table 6: Comparison of Posttest Results
Start and End of the Posttest Period

	Other Customers	**Best Customers**
Start of Posttest Period	0.122^{**} (0.037)	-0.094^{**} (0.019)
End of Posttest Period	0.073^{*} (0.037)	0.016 (0.018)
Complete Posttest Period	0.097^{**} (0.026)	-0.037^{**} (0.013)
Sample Sizes	9,548	9,825

The table reports the *Treatment* variable coefficients when estimating Equation 11 using data from the start and end of the Posttest period. Complete findings (including the omitted coefficients) are reported in Table A6 in the Appendix. Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$. *Significantly different from zero, $p < 0.05$.

The negative Posttest outcome for the *Best* customers is concentrated at the start of the 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 inter-temporal substitution. In studies of inter-temporal substitution in the pricing literature we see a similar pattern, with the "post-promotion 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 Treatment condition leads to a significant increase in demand throughout the Posttest period. Although the estimated effect-size drops from 12.2% to 7.3% 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 coefficient reported in Table 5 for the Total period may underestimate the true size of the effect.

The findings in Tables 5 and 6 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 Treatment 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.

In Section 1, we argued that the substitution interpretation for the negative long-run effect observed among the *Best* customers also has a cross-channel analogy. An implication is that if our interpretation of the results is correct, we should observe a similar effect across channels. Mail and telephone are the primary ordering channels for catalog orders, while alternative channels include both traditional retail stores and Internet stores. Recall that we received demand data for purchases made through both the catalog (mail and telephone) and the company's Internet website. 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 treatment condition led to substitution from the Internet to the catalog channel.

4.3 Cross-Channel Substitution

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 11 separately using both of these dependent measures. The findings are reported in Table 7. Again, for ease of presentation we only report the *Treatment* coefficients (the complete model is reported in Table A7 in the Appendix).

Table 7: Comparison of Test Period Results By Channel

	Other Customers	Best Customers
Catalog Channel	0.116^{**} (0.020)	0.063^{**} (0.011)
Internet Channel	0.303^{**} (0.055)	-0.096^{*} (0.038)
Both Channels	0.138^{**} (0.019)	0.051^{**} (0.010)
Sample Sizes	9,548	9,825

The table reports the *Treatment* variable coefficients when estimating Equation 11 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 Table A7 in the Appendix. Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$. *Significantly different from zero, $p < 0.05$.

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 measure of the total change in Test period demand (in Table 6) may understate the effect; that is, it could be positively biased for the *Best* customers and negatively biased for the *Other* customers.

4.4 Sending Catalogs to Their Best Customers

As we discussed, most companies design their catalog mailing policies by varying their mailing policies and evaluating only the orders received from those same catalogs. This myopic focus on the short-run catalog demand ignores the externalities in other channels and in future periods. For example, the findings in Table 5 indicate that among *Best* customers the short-run response to advertising overstates the long-run response to advertising by a factor of three (5.1% vs. 1.6%). As a result, firms that rely on the short-run response are likely to overinvest in advertising.

To illustrate the implications of this on firm profit, we summarize the profits earned in each condition in Table 8. The profits are calculated as the sum of the items ordered by each customer, multiplied by the profit margin on each item, less 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 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 inter-temporal and cross-channel outcomes we see that the profit result is reversed. The company actually earned a higher average profit in the Control 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.

Table 8: Average Profit Earned Per Customer
Treatment vs. Control

	Control	Treatment	Difference
Best Customers			
Catalog profit during test period	$89.98	$91.56	$1.58
Total test period (including Internet)	$98.74	$100.27	$1.53
Total profit (including Posttest and Internet)	$164.57	$163.84	-$0.73
Sample size	4,921	4,904	
Other Customers			
Catalog profit during test period	$15.50	$15.86	$0.36
Total test period (including Internet)	$19.46	$20.54	$1.08
Total profit (including Posttest and Internet)	$35.06	$37.49	$2.43
Sample size	4,790	4,758	

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.

This interpretation of the findings raises the question as to why companies ignore these long-run and cross-channel effects. We offer to 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 pre-printed 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 website, 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 future purchases to past mailing decisions is even more difficult.

Furthermore, when future purchases are linked to past mailings as part of a controlled experiment, it turns out to be important to consider different customer subsets separately. If the *Best* and *Other* customers are pooled, 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, however. Instead, it reflects the negative effects on the *Best* customers canceling out the continuing positive contributions for the *Other* customers. This could have been easily 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 straight-forward 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 long a straight-forward 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 requiring an infinite series of such tests, while evaluating the long-run impact of these tests requires that companies wait for the long-run to occur.

At least one important question remains. In our model of the relationship between advertising and future demand we assumed that the relationship between advertising and preferences is positive but did not offer an explanation for this relationship. The literature suggests two competing explanations: *information* and *persuasion*. Our ability to compare how the long-run outcome varied across different samples of customers provides an opportunity to distinguish between these two explanations.

4.5 Information or Persuasion?

Recall that the catalogs used in this study contain information about what products are available and the current prices of those products. Yet prices and product availability change quickly in the women's clothing category, and so the incremental information that customers in the Treatment condition received about product availability and prices is unlikely to have significantly affected their future demand. With few exceptions, this information would no longer have been relevant once the Posttest period started (almost four months after the last manipulation).

It is tempting to conclude that the incremental advertising sent to customers in the Treatment condition therefore served a persuasive role. However, before doing so, it is important to also investigate whether the incremental catalogs may have provided other types of information (other than prices and product availability). Recall that the catalogs used in this study also contain descriptions of the company's warranties, ordering procedures, and payment methods, together with implicit cues and explicit claims that may provide customers with information about product quality. It is possible that by exposing customers to this static information, the additional catalog advertising served an informational role that may have influenced future demand.

A recent study by Ackerberg (2001) suggests an approach for distinguishing between these informative and persuasive interpretations. In a study of supermarket demand for yogurt, Ackerberg argues that the informative role of advertising should only affect the behavior of customers who have a need for information. In particular, the effect should be limited to customers who have relatively little experience with the company. Using historical data from a panel of households, he shows that inexperienced customers demonstrate a stronger reaction to television advertising than experienced customers, and concludes that this is evidence that the advertising is performing an informative role. We can use a similar approach to distinguish between the persuasive and informative explanations for the long-run findings in this paper.

Ackerberg argued that the informative role of advertising should only affect the behavior of customers who have a need for information. There should be an analogous moderating effect for the persuasive role of advertising: customers who are all already convinced about the merits of the company should not be susceptible to additional persuasion. This reasoning leads to the two-dimensional interaction summarized in Table 9, predicting which customers are susceptible to additional information and/or persuasion.

Table 9: Identifying Which Customers Are Susceptible To Additional Information and/or Persuasion

	Persuaded Customers	**Unpersuaded Customers**
Informed Customers		Persuasive Role
Uninformed Customers	Informative Role	Informative Role and Persuasive Role

Operationalizing this interaction requires that we identify a measure to distinguish customers who are informed from those who are uninformed. As we discussed, the information at issue is unlikely to be information about current prices and/or product availability. Instead, the relevant information is more likely to concern static characteristics of the company and its products and policies. The experimental manipulation involved mailing additional catalogs and so a good measure of how much prior information customers had about these static characteristics is how many catalogs the customers had received prior to the start of the study. Customers who had already received a lot of catalogs should already have a lot of prior information about the company. If the catalog advertising acts to inform customers about the static characteristics of this company we should not see a response from these customers. Therefore, we use the total number of catalogs received in the five years prior to the start of the experimental manipulation (*Catalogs Received*) as a measure of the degree to which customers are informed.

We also need a measure to distinguish customers who at the time of the test are already persuaded about the merits of the company from those who are not yet persuaded.

Customers who are convinced about the merits of the company are more likely to have recently purchased a relatively large number of items. Therefore, as a measure of prior persuasion we use the total expenditure in the 24-months prior to the start of the experimental manipulation (*Prior Expenditure*).

Before presenting any analysis it is appropriate to consider the sources of variation in the prior mailing policies and customers' prior purchasing levels. Our approach for distinguishing between the persuasion and information explanations requires that there is independent variation between the *Catalogs Received* and *Prior Expenditure* measures. The correlation between the two measures is 0.325 (significantly different from zero, $p<0.01$). This correlation is consistent with our analysis of the impact of catalog advertising on short-run demand: customers who have received more catalogs are more likely to have purchased (see Table 4). However, further investigation reveals that there are two segments of customers who lower the correlation and contribute to independent variation in the measures. We can illustrate these two segments by using the medians of the *Catalogs Received* and *Prior Expenditure* measures to split customers into four segments (see Table 10).

Approximately a third of the customers have high prior expenditure and have received many catalogs, while another third have relatively low prior expenditure and have received fewer catalogs. The large number of customers in these two segments is consistent with a causal relationship between historical mailing decisions and historical demand. Yet we also see approximately 3,000 customers in each of the off-diagonal cells. The 3,056 customers with high expenditure in the previous two years but few catalogs received are typically customers whose first purchase from the company was relatively recent. A recent first purchase means that the company has had relatively few opportunities to send catalogs to this customer. Although the volume of their recent purchases indicates that they have favorable perceptions of the company, the customers have received relatively little information from the company's catalogs.

Table 10: Samples Sizes that Result From Median Splits of the Catalogs Received and Prior Expenditure Measures

	High Prior Expenditure	Low Prior Expenditure
Many Catalogs Received	6,739	3,050
Few Catalogs Received	3,056	6,528

The table reports the sample sizes that result when splitting the sample into sub-samples using a median split of both the *Prior Expenditure* and *Catalogs Received* variables.

The 3,050 customers who have received many prior catalogs but have made few recent purchases are generally customers whose first purchase was made well before the start of the manipulation period. Recall that the *Catalogs Received* measure considers all catalogs mailed in the five years prior to the manipulation period, while the *Prior Expenditure* measure considers purchases within two years of the manipulations. Customers who purchased between two and five years before the test will have received many catalogs but may have made few recent purchases. Although these customers will be well-informed about the static characteristics of the company, their lack of recent purchases suggests that there is an opportunity to raise their persuasion levels.

To estimate the moderating role played by prior information and prior persuasion we modified Equation 11 to incorporate interactions between the *Treatment* variable and both *Catalogs Received* and *Prior Expenditure*:

$$\begin{aligned}\boldsymbol{\beta}\mathbf{X}_i = {} & \beta_0 + \beta_1 \log(Recency_i) + \beta_2 \log(Frequency_i) + \beta_3 \log(Monetary\ Value_i) \\ & + \beta_4 Catalogs\ Received_i + \beta_5 Prior\ Expenditure_i + \beta_6 Treatment_i \\ & + \beta_7 Treatment_i * Catalogs\ Received_i + \beta_8 Treatment_i * Prior\ Expenditure_i\end{aligned} \quad (12)$$

Under this specification the β_7 and β_8 coefficients estimate the moderating influence of the two measures on the long-run impact of the treatment. We report these findings in Table 11, where we also report three benchmark models that include each of these interactions separately and neither of the interactions. In each of the models we pool data from both the *Best* and *Other* customer samples. We again remind readers that this analysis preserves the experimental control due to the random assignment of customers to the two experimental conditions.

Table 11: The Impact of Additional Catalog Advertising on Posttest Demand The Moderating Role of Catalogs Received and Prior Expenditure

	Model 1	**Model 2**	**Model 3**	**Model 4**
Intercept	-1.747** (0.114)	-1.789** (0.115)	-1.731** (0.115)	-1.749** (0.116)
Recency	-0.183** (0.004)	-0.183** (0.004)	-0.183** (0.004)	-0.183** (0.004)
Frequency	0.617** (0.010)	0.617** (0.010)	0.602** (0.010)	0.602** (0.011)
Monetary Value	0.266** (0.025)	0.264** (0.025)	0.257** (0.025)	0.257** (0.025)
Catalogs Received	-0.031* (0.014)	-0.002 (0.018)	-0.026 (0.014)	-0.013 (0.018)
Prior Expenditure	0.010** (0.0004)	0.011** (0.0004)	0.015** (0.0006)	0.015** (0.0006)
Treatment	-0.025* (0.012)	0.075* (0.037)	0.060** (0.014)	0.100** (0.037)
Treatment * Catalogs Received		-0.058** (0.020)		-0.024 (0.020)
Treatment * Prior Expenditure			-0.007** (0.0007)	-0.007** (0.0007)
Log Likelihood	-42,046	-42,042	-41,993	-41,992
Sample Size	19,373	19,373	19,373	19,373

The table reports the coefficient estimates that result from estimating Equation 12 on Posttest demand. Standard errors are in parentheses. **Significantly different from zero, $p < 0.01$. *Significantly different from zero, $p < 0.05$.

There are several findings of interest. First, the interaction between the *Treatment* effect and *Prior Expenditure* (in Models 3 and 4) is negative and highly significant. This indicates that the long-run impact of the experimental manipulation was moderated by the level of customers' recent prior expenditure. The favorable long-run outcome was limited to customers with relatively low levels of prior expenditure; a finding which is consistent with the earlier contrasting results for the *Best* and *Other* customer samples.

Second, the interaction between the *Treatment* effect and *Catalogs Received* is also negative. However, this coefficient is only significant in Model 2, where the interaction with *Prior Expenditure* is omitted. It is possible that *Prior Expenditure* acts as a proxy for *Catalogs Received* in Model 2. When both interaction terms are included (Model 4) the interaction between the *Treatment* and *Catalogs Received* is no longer significant. A likelihood ratio test between Models 3 and 4 indicates that the addition of the *Catalogs Received* interaction term does not increase the explanatory power of the model. We conclude that the findings suggest that the number of prior *Catalogs Received* does not independently moderate the treatment effect.

Finally, in the absence of both interaction terms (Model 1) we see that the coefficient for the *Treatment* variable is negative and significant. This contrasts with the findings in Models 2, 3 and 4, and highlights the difficulty of analyzing the long-run response to advertising. Failure to anticipate the heterogeneity in the long-run response would have led to the (erroneous) conclusion that the intervention had a negative long-run impact on all customers.

We conclude that the findings in Table 11 offer support for the persuasive view. The expansion in long-run demand does not appear to apply to the company's most valuable customers. Intuitively, it is hard to make the company's best customers any better as these customers are already using the company to satisfy their category needs - there is a limit to how many clothes even the most loyal customer can wear!

5. 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 current advertising can impact future demand. Interestingly, the impact is quite heterogeneous. 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 substitution.

In contrast, among the less valuable customers, who had purchased less frequently and/or less recently, advertising had a positive impact on future demand. We note that these customers are also susceptible to temporal substitution, and so the favorable long-run outcome occurred despite the adverse effects of temporal substitution (if any).

We investigated two alternative explanations for this long-run outcome. The first explanation focused on the persuasive role of advertising and the second focused on advertising's informative role. To distinguish these explanations we took advantage of predicted heterogeneity in how sensitive different types of customers are to additional information and/or additional persuasion. The findings offer support for the persuasive argument. Customers most affected by the advertising were those whose past behavior indicated that there was an opportunity to increase their consumption through additional persuasion.

The findings also 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 over-mail 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.

Our findings help untangle the questions about advertisings effects on long run demand. It turns out that advertising causes both increases and decreases in future demand, depending on the type of customer. We can identify which types of customers are likely to be in each group. 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 business practice.

6. References

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

Ackerberg, D. A. (2001), "Empirically Distinguishing Informative and Prestige Effects of Advertising," *Rand Journal of Economics*, 32(2), pp 316-33.

Ackerberg, D. A. (2003) "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination", *International Economic Review*, 44(3), August, 1007-40.

Advertising Age (2005), *Fact Pack: 3rd Annual Guide to Advertising and Marketing*, February 28, Crain Communications, Inc.

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

Bagwell, K. (2005), "The Economic Analysis of Advertising," forthcoming in Mark Armstrong and Rob Porter (eds.), *Handbook of Industrial Organization*, 3, North-Holland: Amsterdam.

Becker G. S. and K. M. Murphy (1993), "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108, 942-964.

Benham, L. (1972), "The Effects of Advertising on the Price of Eyeglasses," *Journal of Law and Economics*, 15, 337-352.

Blattberg, R. and S. Neslin (1990), *Sales Promotions*, Prentice Hall.

Borden, N. H. (1942), *The Economic Effects of Advertising*, Richard D. Irwin.

Comanor, W. S. and T. A. Wilson (1967), "Advertising, Market Structure and Performance," *The Review of Economics and Statistics*, 49, pp. 423-40.

Comanor, W. S. and T. A. Wilson (1974), *Advertising and Market Power*, Harvard University Press.

Direct Marketing Association (2003), *The DMA 2003 Statistical Fact Book*, New Jersey: DMA.

Direct Marketing Association (2005), *The DMA 2005 Statistical Fact Book*, New Jersey: DMA.

Elsner R, M. Krafft, and A. Huchzermeier (2003). "Optimizing Rhenania's Mail-order Business through Dynamic Multilevel Modeling," *Interfaces*, Vol. 33(1), 50-66.

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

Gönül, F. and M. Shi (1998). "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science*, 44(9) 1249-62.

Grossman, G. and C. Shapiro (1984), "Informative Advertising with Differentiated Products," *Review of Economic Studies*, 51, 63-81.

Hendel, I. and A. Nevo (2002), "Sales and Consumer Inventory," NBER working paper 9048, NBER Cambridge MA.

Hendel, I. and A. Nevo (2003), "The Post Promotion Dip Puzzle: What do the Data Have to Say?" *Quantitative Marketing and Economics*, 1(4), 409-424.

Hendel, I. and A. Nevo (2005), "Measuring the Implications of Sales and Consumer Inventory Behavior," NBER working paper 11307, NBER Cambridge MA.

Ippolito P. M. and A. D. Mathios (1990), "Information, Advertising and Health Choices: A Study of the Cereal Market," *RAND Journal of Economics*, 21, 459-480.

Kaldor N. (1950), "The Economic Aspect of Advertising," *Review of Economic Studies* 18 (1), 1-27.

Kihlstrom R. and M. Riordan (1984), "Advertising as a Signal," *Journal of Political Economy*, 92, 427-450.

Krishnamurthi, L. and S. P. Raj (1985), "The Effect of Advertising on Consumer Price Sensitivity," *Journal of Marketing Research*, 22, May, 119-29.

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

Lodish, L.M., M. M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, and M. E. Stevens (1995a), "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments," *Journal of Marketing Research*, 32(2), 125-139.

Lodish, L. M., M. M. Abraham, J. Livelsberger, B. Lubetkin, B. Richardson, and M. E. Stevens (1995b), "A Summary of Fifty-Five In-Market Experimental Estimates of the Long-Term Effect of TV Advertising," *Marketing Science*, 14(3), part 2, G133-G140.

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

Milyo, J. and J. Waldfogel (1999), "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart," *American Economic Review*, 89 (5), 1081-1096.

Nelson, P. (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78, May 311-29.

Nelson, P. (1974), "Advertising as Information," *Journal of Political Economy*, 82, August, 729-54.

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

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

Schmalensee, R. (1978), "A Model of Advertising and Product Quality," *Journal of Political Economy*, 86 (3), 485-503.

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

Simester, D. I., Sun P. and J. N. Tsitsiklis (2005), "Optimal Catalog Mailing Policies," *Management Science*, forthcoming.

Stigler, G. J. (1961), "The Economics of Information," *Journal of Political Economy*, 69, June, 213-25.

Telser, L. G. (1964), "Advertising and Competition," *Journal of Political Economy*, 72(6), December, 537-62.

Table A5: Comparison of Test Period, Posttest Period and Total Results

	Posttest Period		Total Period	
	Other Customers	Best Customers	Other Customers	Best Customers
Intercept	-0.872** (0.168)	-3.195** (0.146)	-0.432 (0.105)	-3.168** (0.092)
Recency	-0.289** (0.008)	-0.146** (0.005)	-0.281** (0.005)	-0.137** (0.003)
Frequency	0.465** (0.013)	0.723** (0.010)	0.480** (0.008)	0.738** (0.006)
Monetary Value	0.221** (0.037)	0.515** (0.032)	0.350** (0.023)	0.705** (0.020)
Treatment	0.097** (0.026)	-0.037** (0.013)	0.124** (0.015)	0.016* (0.008)
Log Likelihood	-13,145	-28,891	-25,422	-44,298
Sample Size	9,704	9,834	9,704	9,834

The Posttest findings reports the coefficients from Equation 11 estimated using data from the Posttest period. The Total Period findings report the coefficients from Equation 11 estimated using data from the entire period (Test and Posttest). Standard errors are in parentheses.
**Significantly different from zero, $p < 0.01$.
*Significantly different from zero, $p < 0.05$.

Table A6: 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
Intercept	-2.200** (0.256)	-4.656** (0.209)	-1.033** (0.217)	-3.172** (0.202)
Recency	-0.272** (0.012)	-0.162** (0.008)	-0.305** (0.011)	-0.131** (0.007)
Frequency	0.466** (0.019)	0.754** (0.014)	0.467** (0.019)	0.692** (0.013)
Monetary Value	0.345** (0.056)	0.688** (0.047)	0.117** (0.048)	0.354** (0.045)
Treatment	0.122** (0.037)	-0.094** (0.019)	0.073* (0.037)	0.016 (0.018)
Log Likelihood	-8,092	-19,387	-8,254	-20,447
Sample Size	9,704	9,834	9,704	9,834

The Internet Channel findings reports the coefficients from Equation 11 estimated using purchases from the start and end of the Posttest period. Standard errors are in parentheses.
**Significantly different from zero, $p < 0.01$.
*Significantly different from zero, $p < 0.05$.

Table A7: Comparison of Test Period Results By Channel

	Internet Channel		Catalog Channel	
	Other Customers	Best Customers	Other Customers	Best Customers
Intercept	-2.460** (0.343)	-9.563** (0.422)	-1.491** (0.144)	-4.110** (0.122)
Recency	-0.447** (0.016)	-0.066** (0.016)	-0.249** (0.007)	-0.136** (0.004)
Frequency	0.598** (0.028)	0.829** (0.028)	0.473** (0.011)	0.742** (0.008)
Monetary Value	0.326** (0.075)	1.362** (0.093)	0.438** (0.031)	0.783** (0.027)
Treatment	0.303** (0.055)	-0.096* (0.038)	0.116** (0.020)	0.063** (0.011)
Log Likelihood	-4,739	-9,623	-17,882	-32,568
Sample Size	9,704	9,834	9,704	9,834

The Internet Channel findings reports the coefficients from Equation 11 estimated using purchases through the Internet channel in the Test period. The Catalog Channel findings use purchases through the Catalog channel.
**Significantly different from zero, $p < 0.01$.
*Significantly different from zero, $p < 0.05$.