# 1. Introduction

When prospecting for new customers firms are often forced to target customers who did not respond in the past. This is because newly acquired customers are removed from the target population: unless the population renews, the remaining prospects are those who failed to respond to every previous promotion. What makes targeting these remaining prospects profitable is that some will respond to future promotions, despite not responding in the past. This could occur for a variety of reasons including: (a) promotions changing over time, (b) customers overlooking the prior promotions, (c) customer preferences changing, or (d) the likelihood of a response increasing with repeated exposure.

We can illustrate the issue by describing the problem faced by the firm that provided the data for this study. It is a retailer selling mass merchandise that recruits new customers as members using direct mail. The prospect pool is comprised of households living close to a store, excluding any households that are already members. The retailer has already sent multiple promotions to many of these households. The challenge for the firm is to decide how to target this increasingly pure pool of non-responders.

The paper proposes an approach for using decisions by past responders to help firms target non-responders. The proposed approach is motivated by the insight that some regions are exhausted of future responders, while other regions contain potential responders who may just require additional exposures. The possibility that households may require additional exposures before they will respond is a basic tenet of media planning, and has given rise to the terms “wearin” and “wearout”. An advertisement is said to be *wearing in* in a region if additional exposures increase the probability of a response in that region. Alternatively, an advertisement is said to have *worn out* in a region once additional exposures no longer increase the probability of a response in that region (see Pechmann and Stewart 1988).

A key insight in this paper is that wearin and wearout have different implications for the timing of responses when multiple promotions are sent. For example, if two waves of the same promotion are mailed to the same households, and the households are increasingly *wornout*, then we would expect a larger response to the first wave of promotions than to the second one. In contrast, if the promotion is still wearing in with these households, then we may see a larger response to the second wave than the first one. This, in turn, may have implications for the likelihood that other households will respond in future waves. We might expect a larger response to future mailings if there is an increasing response rate between Waves 1 and 2 compared to when the response rate is decreasing (even if the total response in Waves 1 and 2 is held constant). This reasoning suggests that the slope of the response rate between Waves 1 and 2

can help to predict which regions will have a larger response to future mailings, and therefore which regions to continue to target.

In this paper we present findings supporting this relationship. The findings are obtained using data from a large-scale field experiment in which a sequence of promotions are mailed to the same pool of prospective customers. The experimental design includes both a pre-treatment period and a treatment period.

**Pre-Treatment Period**

In fall 2015 the firm mailed a promotion to almost 400,000 households. These households were all prospective customers, who were not customers of the retailer at the time. The promotion was designed to induce the households to become customers. Identical mailings were sent in two waves, and all 400,000 households received both of these mailings. The response to each of these waves provides data to measure the slope of the response rate (between Waves 1 and 2) in each region. More details are provided in Section 3.

**Treatment Period**

In spring 2016 we mailed the same promotion to a randomly selected set of approximately 87,000 prospective customers. These households were randomly selected from the 400,000 households that received the fall mailings and did not respond in the fall (additional details are provided in Section 3). This design allows us to measure the slope of the response in fall in each region using a large sample of customers (all 400,000). We then use this slope to predict the response to subsequent mailings in spring by a smaller randomly selected subset of customers.

An important difference between fall 2015 and spring 2016 is that any household that responded in fall was removed from the spring mailing pool. After responding in fall, these customers become existing customers and are no longer prospects. The timing of the responses by these fall responders is used to decide which of the 87,000 prospects to mail to in spring. It is in this respect that we use responders to help target non-responders.

We experimentally varied the sequences of promotions mailed in spring, including how frequently the promotion was mailed. This allows us to validate the robustness of the findings. The 87,000 households in the spring prospect pool were randomly assigned into five experimental conditions. In one experimental condition households were mailed the promotion three times. In three other conditions households were mailed twice or just once. The fifth experimental condition was a control condition, in which households were not mailed at all.

Our analysis includes two types of evaluations. First, we calculate the lift in response rates attributable to each treatment and investigate whether the timing of the pre-treatment responses helps to explain its variation. To calculate the lift in (treatment period) response rates,

we subtract the response rate in the control condition that received no mailing from the response rate in each of the four treatment conditions.

Second, we construct a targeting model to decide which households to mail to. In particular, we use the spring outcomes in a subset of the regions to develop the targeting model, and then use the remaining regions as a holdout sample with which to evaluate the targeting model. We investigate whether including variables describing the timing of the pre-treatment responses improves the performance of the targeting model.

**Measuring the Timing of the Response in the Pre-Treatment Period**

Our intuition that the slope of the response to previous mailings can predict the outcome of future mailings relies on reasoning that in some regions a promotion may still be wearing in, while in other regions it is wearing out. To test this intuition and use it to improve targeting decisions, we need a measure for distinguishing regions in which a promotion is wearing in versus wearing out.

Measuring the timing of past responses is challenging for two reasons. First, timing information is complicated as different customers respond at different times. We need to find measures that can efficiently summarize this information. Second, the timing information is also very sparse. We only measure the behavior of households that responded, and for a prospecting campaign the proportion that responds is typically very low. The DMA reports average prospecting response rates of approximately 6 per 1000 (0.6%) households mailed (DMA 2015).

We propose three measures for the timing of the response in the pre-treatment period. The first two measures are relatively simple, while the third measure is more sophisticated.

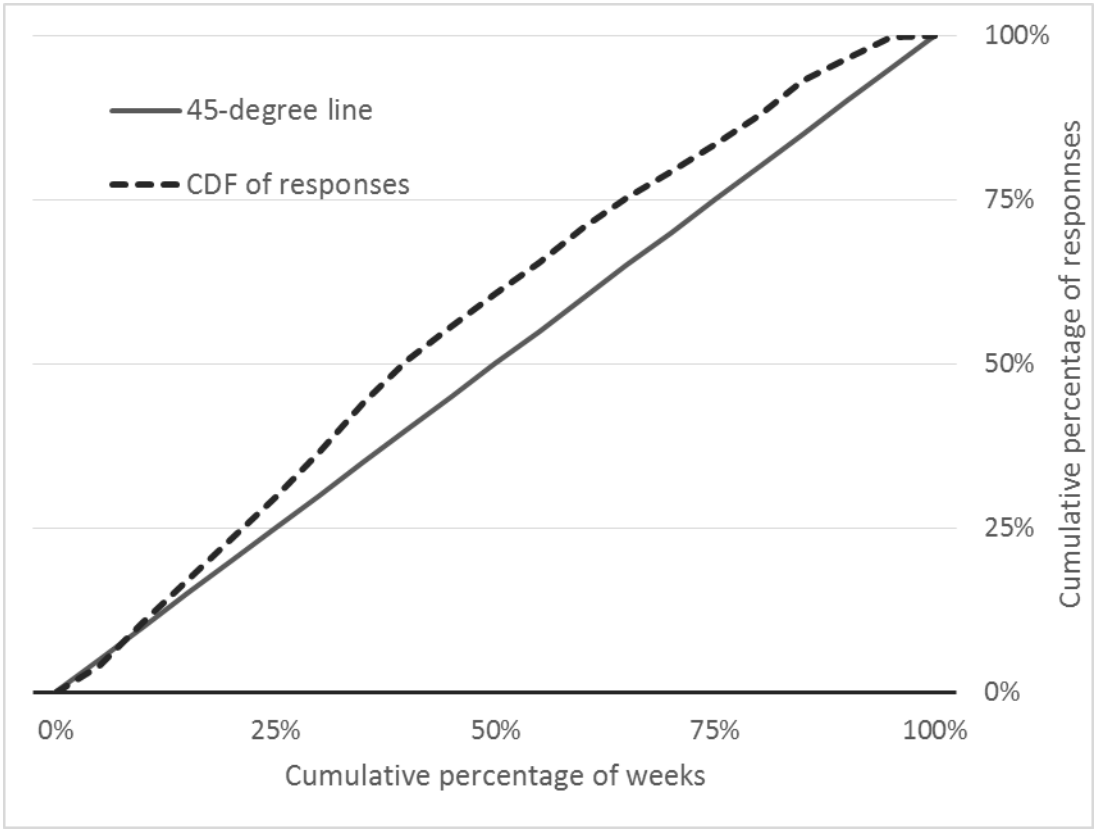
The first metric measures the share of responses received before the Wave 2 mailing date (*Response Before Wave 2*). The second calculates the *Average Response Date* in each geographic region. This is the number of days between the Wave 1 mailing date and a household's response date, averaged across the households that responded in that region.

Our third, more sophisticated measure, starts by constructing an empirical CDF in each region of the fall pre-treatment response date distribution. The support of the CDF represents the cumulative proportion of weeks, starting at the Wave 1 mailing date and finishing at the end of the measurement period. The CDF describes the cumulative proportion of all responses received in that region that had occurred by the corresponding week. As an illustration, we report the aggregate CDF in Figure 1 using the responses to the fall pre-treatment mailings (from all regions).

We can more easily interpret the CDF by comparing it to the 45-degree line. The difference between the CDF and the 45-degree line provides a measure of the timing of the fall responses. If the CDF lies above the 45-degree line, this indicates that response occurred relatively early in

the measurement period. In contrast, if the CDF lies below the 45-degree line, then the responses occurred relatively late.[1]

**Figure 1. CDF of Aggregate Fall Responses by Week**




This figure reports the CDF of responses to the fall pre-treatment mailings aggregated across regions. The x-axis identifies the percentage of weeks and the y-axis identifies the percentage of responses.

Recall that the illustration in Figure 1 aggregates data across all regions. Our proposed approach to measuring timing constructs a separate CDF for each region using the fall pre-treatment responses. These empirical CDFs provide a rich measure of the timing of the fall responses per week. We use the distances between these empirical CDFs as an input to a fuzzy clustering algorithm. Rather than attributing regions to a specific cluster, the fuzzy clustering algorithm yields soft clusters, based on the degree of membership of a region to each cluster. Degree of membership to a cluster indicates the extent to which the timing of responses in a region

[1] An analogous approach is used to construct the Lorenz curve, which economists have used to measure inequality in income and wealth distribution (Lorenz 1905, Gini 1955). Brynjolfsson, Hu, and Simester (2011) also use a similar approach to measure inequality in the distribution of sales on the Internet. The area between the 45-degree line and the CDF of product sales provides a measure of the long tail of sales on the Internet.

followed a similar distribution to other regions in that cluster. As we shall see, the timing of the responses in each cluster yields a relatively clear depiction of the shape of the response curves in these clusters. We use the degrees of membership of the regions to the clusters (*Membership to Clusters*) as our third timing measure. We can interpret this measure as the degree to which the timing of the fall pre-treatment responses is similar to that of the cluster of regions exhibiting wearin (versus the cluster of regions exhibiting wearout).

**Summary of Key Findings**

There are several findings of interest. First, there is robust evidence that regions in which the response rate increased between pre-treatment waves in fall, are also regions in which the response was higher in the spring treatment period. This is consistent with our interpretation that an increasing response rate is an indication that a promotion is still wearing in. It is these regions in which we expect additional mailings to yield more incremental responses. In contrast, if the response rate is decreasing between the two waves in the fall, this suggests that the promotion is wearing out, and so additional repetition in spring is likely to yield fewer incremental responses.

Our findings also confirm that the information that the timing of past responses provides about future response rates can also improve the accuracy of a targeting model. We compare targeting models that include versus omit the timing measures. We show that including the timing measures results in more accurate mailing decisions, and this in turn leads to higher average profits.

Finally, our findings also allow us to compare the performance of the three proposed timing variables. The *Response Before Wave 2* and *Membership to Clusters* timing variables both yield larger improvements in the performance of the targeting model than the *Average Response Date* measure. The performance of the *Response Before Wave 2* is particularly notable. The ease of constructing this variable makes the option of using the timing of past responses accessible to essentially any firm engaged in targeting prospective customers.

**Structure of the Paper**

The paper proceeds in Section 2 with a review of the literature. In Section 3 we describe the studies and data used in the paper in greater detail. In Section 4 we present preliminary findings, and then in Section 5 we present the results of the targeting models. The paper concludes in Section 6.

## 2. Literature

Our research relates to three streams of literature: targeting methods, the construction of new variables for optimization, and the effects of wearin and wearout in advertising. We discuss each of these areas together with our contribution next.

The marketing literature has long recognized the importance of targeting in a range of marketing contexts. For example, Zhang, Netzer, and Ansari (2014) develop a price targeting model, where the objective function balances short-term profit maximization with long-term investment in relationships. Ascarza (2017) investigates customer churn in the telecommunications industry. Dubé and Misra (2017) combine machine learning and Bayesian approaches for online price targeting. Luo, Andrews, Fang, and Phang (2013) focus on targeting of advertising in a mobile setting. Simester, Sun, and Tsitsiklis (2006) consider a catalog mailing optimization problem that incorporates dynamic effects. Zhang and Wedel (2009) study the effectiveness of customized promotions in online and offline stores, and demonstrate that online and offline targeting are effective at any level of aggregation.

We focus on customer acquisition (prospecting). Targeting of prospective customers introduces a unique set of challenges. Firms generally do not have any prior purchase history for prospective customers and so they are forced to rely upon demographic or geographic data. This data changes only very slowly over time, and so subsequent prospecting models are trained using the same data as previous models. As a result, it is common practice for firms to repeatedly, but not necessarily optimally, mail promotions to the same prospective households that failed to respond to previous promotions.

One way to improve targeting of prospective customers is to provide an additional source of information about which prospects are most likely to respond. Our contribution is an example of this type of improvement. There are many other examples of papers that have sought to improve targeting models by introducing additional information. For example, Rossi, McCulloch, and Allenby (1996) compare improvements in a coupon targeting model when varying how much purchase history is available. They demonstrate significant improvements when using longer purchase histories. Toubia, Goldenberg, and Garcia (2014) demonstrate that social interaction data can improve forecasts of new product market penetration. Goel and Goldstein (2013) demonstrate the predictive power of geographic, behavioral, and demographic variables describing an individual's contacts in a social network. They conclude that in the absence of detailed purchase histories, social interaction data can help to improve predictions of individual behavior.

The targeting variables constructed in this paper are obtained from a field experiment. Running field experiments to train targeting models is costly, introducing a tradeoff between exploration and exploitation. Schwartz, Bradlow, and Fader (2017) investigate this tradeoff in a customer

acquisition setting and propose an approach for efficient online experimentation using display advertising. With better experimentation, the proposed model achieves 8% higher customer acquisition rates without any additional costs to the firm. Li et al. (2015) investigate the number of experiments required to estimate a large matrix of cross-product price elasticities. They discuss how their findings generalize to marketing decisions other than pricing, such as selecting which products to promote.

We also use data from the field experiment for validation.[2] Field experiments have been recognized as the gold standard for validating targeting models (see Simester 2017 for a review). For example, Dubé, Fang, Fong, and Luo (2017) conduct a field experiment to validate a price targeting mechanism. The advantage of using field experiments for validation is that the data can provide a "model free" evaluation of the competing models. In particular, candidate models can be compared without reliance on the functional form or modeling assumptions in any of the models.

Our analysis focuses on the wearin and wearout effects of repeated promotions.[3] We are not the first to formally model wearin and wearout effects. Naik, Mantrala, and Sawyer (1998) estimate a dynamic model of the effect of promotion repetition on customer awareness. Their model shows that pulsing can leverage wearin and wearout effects and generate greater total awareness than continuous advertising. Bass, Bruce, Majumdar, and Murthi (2007) develop a dynamic Bayesian model of the advertising-sales relationship for multiple advertising messages. Their model quantifies wearout of each message.

Since as long as 1971 it has been recognized that the rate at which a promotion wears in or out varies across customers (Ray and Sawyer 1971). This process depends upon a variety of factors, including a customer's motivation and ability to process an advertisement, together with exposure to competing brands (Pechmann and Stewart 1988). We incorporate wearin and wearout measures in a reduced-form model by measuring how response rates change over time. This can be interpreted as augmenting demographic data with an additional measure of customer heterogeneity. We show that this can improve targeting of prospective customers.

[2] Because the field experiment also provides training data, we validate the candidate models using cross-validation.

[3] Recall from the introduction that an advertisement is said to be "wearing in" in a region if additional exposures increase the probability of a response in that region. Alternatively, an advertisement is said to have "worn out" in a region once additional exposures no longer increase the probability of a response in that region (see Pechmann and Stewart 1988).

## 3. Description of the Data

The data in this study was provided by a mass merchandise retailer. The retailer sells a broad range of products including perishables, sundries, and durables. When customers purchase at this retailer they must identify themselves with their membership cards, which allows us to link each transaction to each customer.

The retailer prospects for new members by sending promotions by mail to households that live in the geographic proximity of its stores. The households are identified by purchasing a mailing list from a third party vendor. The retailer identifies "regions" using USPS carrier routes, which represent the routes that each letter carrier is responsible for. On average a carrier route comprises approximately 400 households. The retailer's targeting problem is to decide which carrier routes to mail promotions to, and how frequently to send them. The firm mails the same promotion to every household in a carrier route, after excluding households that already have memberships to the store.[4] This filtering of past responders ensures that the remaining set of households becomes an increasingly pure pool of non-responders.

The retailer provided ten descriptive variables describing each carrier route. Some of these variables (such as distance to the nearest store) are constructed by the retailer, others (such as household income) are provided by the third party vendor that supplies the mailing list. These variables vary at either the carrier route level or the 5-digit zip code level (where a zip code is a collection of carrier routes). The full list of variables, together with their definitions and descriptive statistics, are provided in the Appendix.

We next describe in more detail the fall 2015 pre-treatment mailings and then review the design of the spring 2016 mailing treatments.

**Fall 2015 Pre-Treatment Period**

The pre-treatment activities were designed to measure the extent to which a promotion was wearing in or wearing out in each region. In particular, the firm sent two waves of an identical promotion to prospective customers in each region, and measured whether the response to these promotional offers increased or decreased between the two waves.

We restricted attention to carrier routes with at least 200 households and randomly selected 796 carrier routes from the universe of carrier routes located in the neighborhood of the retailer's stores. In total the fall pre-treatment mailings were sent to 388,686 households distributed

[4] The USPS charges a higher postage if the firm sends different mailings to different households in the same carrier route. The firm also has only relatively weak measures with which to distinguish prospective households within a carrier route. Notably, the firm does not have past purchasing data for these prospective customers. Past testing has also revealed that purchasing household level demographic data from third parties does not yield improvements in targeting performance compared to purchasing carrier route level measures. This is presumably due to noise in the household level demographic measures.

across these 796 carrier routes. All of the households in a carrier route received the fall pre-treatment mailing (except existing customers, which were filtered out).

The promotional offer was a trial membership to the store. The free trial enabled customers to shop in the store for a limited period. If a customer chose to remain a member after the trial period, she needed to purchase a membership at the full price. While the firm had sent promotions for free-trial memberships in previous years, this promotion was unusual compared to past promotions in that the trial period was longer than past free trial promotions. The promotional offers were mailed to prospective customers using a postcard, which was printed on both sides and highlighted the offer on each side.

The 388,686 households were mailed the free trial promotion twice. We will refer to these two mailings as Wave 1 and Wave 2. The mailings were 37 days apart.[5] The dates are approximate because the "in-home" mailing dates vary due to imprecision in the mail delivery process. This imprecision contributes an additional source of noise in the timing of the responses to the pre-treatment mailings.

**Spring 2016 Treatment Period**

The treatment period occurred in spring 2016, approximately six months after the fall pre-treatment mailings. In the treatment period we randomly assigned customers to receive the same promotional mailing that they received in fall. We then measured the response in each region in spring, and compared it to the timing of the fall responses. This allows us to test our prediction that if the response increased between waves in fall, the promotion was still wearing in. In contrast, if the fall response rate decreased between waves, the promotion was wearing out. For this reason, we expected a higher response in spring in the regions in which the response increased between waves in fall, compared to regions in which it decreased (all else being equal).

The spring experiment involved 87,184 households, which were randomly selected from households in the 796 carrier routes. We filtered out any households that responded to the fall pre-treatments and became members before the spring treatment. On average 110 households in each carrier route were randomly selected to participate in the spring treatments. The other households received different mailings unrelated to this experiment (see the Appendix for a more complete description of these unrelated conditions).

The households were randomly assigned to five treatment groups, yielding sample sizes of approximately 17,400 households in each treatment group (actual sample sizes are reported in Table 1). The randomization was conducted at the household level, and so every carrier route includes five subsets of households assigned to the five treatments. The treatments included up

[5] There was not enough time between waves to filter out households that responded to Wave 1 before mailing Wave 2 (the lead times for mailing decisions were longer than the time between waves). Therefore, a household received both promotions even if it responded to the first promotion.

to three waves of mailing. The promotion was identical in each wave and the mailing waves averaged 32 days apart.[6] The treatments are summarized in Table 1.

The promotional offer used in spring was identical to the fall offer: a free trial membership to the store. In spring the promotional offers were mailed to each household using a letter, while in fall the offers were mailed using a postcard. We note that this difference is common across the spring experimental conditions.

**Table 1. Summary of the Spring Experimental Treatment Conditions**

| Condition | Wave 1 | Wave 2 | Wave 3 | Sample Size |
|---|---|---|---|---|
| Control | No Mailing | No Mailing | No Mailing | 17,420 |
| Trial No No | Trial Promotion | No Mailing | No Mailing | 17,382 |
| No Trial No | No Mailing | Trial Promotion | No Mailing | 17,556 |
| Trial Trial No | Trial Promotion | Trial Promotion | No Mailing | 17,382 |
| Trial Trial Trial | Trial Promotion | Trial Promotion | Trial Promotion | 17,388 |

This experimental design provides an almost ideal way to measure the impact of repetition; the same retailer mails the same promotion to households in the same carrier routes in multiple waves a similar time apart.

In the next section we present initial findings comparing the aggregate response to the spring treatments. Throughout the paper we use the following symbols: $^{**}$ indicates significantly different from zero at $p < 0.01$, $^{*}$ significantly different from zero at $p < 0.05$, $^{\dagger}$ significantly different from zero at $p < 0.10$ (all 2-tail tests).

## 4. Initial Findings

In this initial analysis we present two preliminary sets of findings. First we describe the incremental weekly response in the spring treatment conditions (compared to the control condition). We then introduce the three timing measures and use them to investigate whether the timing of responses to the fall pre-treatment mailings helps to explain variation in the response to the spring treatments.

[6] Wave 2 occurred approximately 21 days after Wave 1. Wave 3 occurred approximately 42 days after Wave 2. The research team had no control over the timing of the mailing waves.

The outcome measure of interest is whether a household became a member during the spring treatment period. We use a binary outcome measure, which is equal to one if the customer responded and zero otherwise. We then average this binary measure across households in a carrier route to generate a carrier route-level average. We will later take into account the profitability of these responses when comparing targeting models (in the next section).

**Aggregate Weekly Response to the Spring Treatments**

In the Appendix we summarize the lift in response rates in the spring treatment conditions, calculated as the difference between the response rate in each treatment condition and the response rate in the control condition.[7] To preserve confidentiality we re-scale the response rates by multiplying them by the same random number.

The findings clearly illustrate the responses attributable to the treatments. The increase in the response rate was similar across three of the treatment conditions. Notably there was only a slightly higher lift in responses in the treatments that had multiple mailings, compared to the Trial No No treatment, which had a single mailing. This evidence that repetition of the mailings yields only a small incremental response suggests that in many of the carrier routes the promotion may be wearing out.

The smallest increase in response was in the No Trial No treatment. This may seem surprising as households in this treatment received the same number of mailings as households in the Trial No No treatment. The first mailing in the No Trial No treatment occurred 21 days after the other treatments, and so we might wonder whether the smaller lift in response in this condition is due to truncation of the period over which we measured responses. We believe that this is unlikely to explain the lower response in this treatment. The firm estimates that the response to a mailing is generally measurable within 13 weeks of the mailing date. This is consistent with the industry standard. For example, the Direct Marketing Association reports that the measurable response to a mailing extends over approximately 13 weeks (DMA 2006). We can also investigate this explanation using the timing of the responses to the No Trial No treatment. In the Appendix we report the incremental response by week (compared to the control) for the No Trial No treatment. The lift in response appears to be fully measured by the 17$^{th}$ week after the first wave mailing date, which is 14 weeks after the second wave mailing date. Instead, the smaller effect in the No Trial No condition compared to the Trial No No condition may be attributable to seasonality; sending a single mailing in Wave 1 was more effective than sending a single mailing three weeks later.

We next introduce the three measures that we will use to describe the timing of the response to the fall pre-treatment mailings.

[7] The average response rate is calculated as the number of households that responded in a condition, divided by the total number of households in that condition.

**The Timing of the Pre-Treatment Responses**

The central question in this paper is whether the timing of the fall responses provides information about which regions have a larger response to the spring treatments. We have conjectured that if in fall the response in a carrier route was higher (lower) to the first wave than the second wave, then the promotion is more likely to be wearing out (wearing in), suggesting the response in spring will be relatively low (high) in that carrier route compared to other carrier routes.

Recall that the first and second pre-treatment mailing waves in fall occurred 37 days apart. This means that the response to the first mailing wave was not complete when the second mailing occurred. As a result, after the $37^{th}$ day it is unclear which mailing wave a customer was responding to. To evaluate whether the response to the first fall mailing wave was higher or lower than the response to the second mailing wave, we cannot simply compare the response to each wave.

We also note that the response curve for the second mailing wave was not the same as the response curve for the first mailing wave. We illustrate this in the Appendix where we report the incremental response curves (by week) for Waves 1 and 2 in spring.[8] The shape of the incremental response to the two mailing waves is not the same. In particular, the incremental response to Wave 2 is concentrated in the first few weeks after the mailing date, and has a much shorter tail.

As we discussed in the introduction, the timing of the pre-treatment responses in fall is a relatively rich type of information. For example, separately measuring the responses in each of the 17 weeks in the response window requires 17 variables. Using simple measures to summarize this timing risks loss of important information. On the other hand, using too many variables introduces a risk of over-fitting, especially given how sparse the data is. For example, while we can construct a parametric model of the aggregate response curves for each fall mailing wave, we do not have enough data to reliably fit response curves for each carrier route. Therefore, we propose a method for capturing this relatively high dimensional information using relatively few variables.

We developed three alternative measures of the timing of the response to the two fall mailing waves. The first approach was to simply measure the proportion of the total number of responses across both waves that were received before the Wave 2 mailing date (*Response Before Wave 2*). This relatively aggregate approach only distinguishes between timing in two pre-treatment sub-periods (the first 37 days versus the next nine weeks).

A second simple timing measure calculates the *Average Response Date* in each carrier route. This is the number of days between the Wave 1 mailing date and a household's response date, averaged across the households that responded in that carrier route. This approach is sensitive

[8] The response curve for Wave 1 is represented by the weekly response in the Trial No No condition. The response curve for Wave 2 is represented by difference in the weekly response in the Trial No No and Trial Trial No conditions.

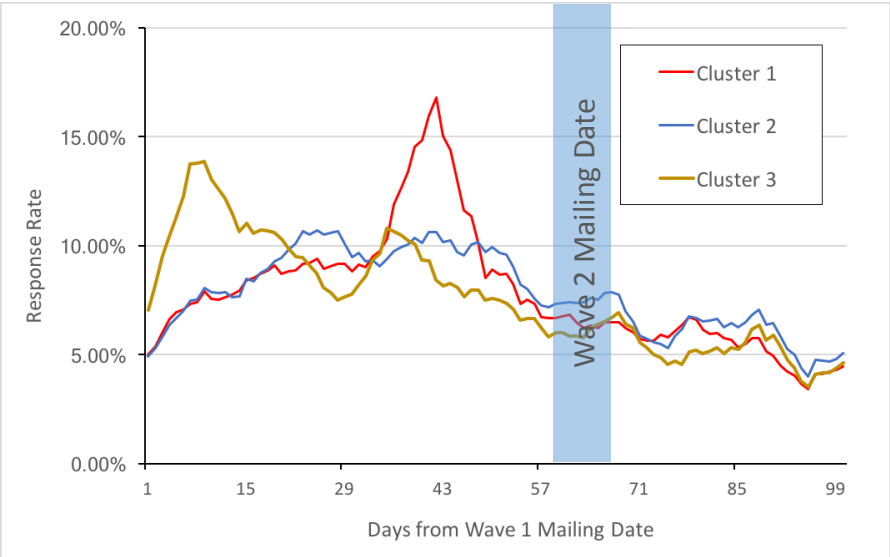
not just to whether the response occurred before versus after the Wave 2 mailing date, but also to timing differences within each sub-period. However, a limitation of the *Average Response Date* is that it only describes the mean of the response timing, and is not sensitive to the variance in this timing.

To construct the third timing measure, we first represent the timing of carrier routes' responses using empirical CDFs. These empirical CDFs are 17-dimensional vectors $\vec{x}^c = \{x_i^c\}_{i=1..17}$, where $x_i^c$ is the share of responses in the carrier route $c$ that occurred by week $i$. We can characterize whether carrier routes exhibit similar response timing patterns by measuring the distance between these vectors.[9] We first calculate distances between the empirical CDFs of the carrier routes and then apply a fuzzy C-means clustering algorithm to identify groups of carrier routes with similar distributions of the response timing (Dunn 1973, Bezdek 1981). Fuzzy C-means is a clustering method that allows observations to belong to two or more clusters with different degrees of membership. We use the degrees of membership as our third timing measure and call the new variables *Membership to Clusters*.

The number of clusters for the fuzzy C-means algorithm is an important decision. Using few clusters leads to loss of information, while using too many clusters leads to redundancy and lack of interpretability. In our study, we use three clusters. Figure 2 reports the average response rates across carrier routes weighted by the degrees of membership to the three clusters. Carrier routes with a high degree of membership to the first cluster are characterized by a high response to the Wave 2 mailing (wearin). The second cluster aggregates carrier routes with low overall response to the mailings, and the third cluster contains carrier routes with high response to Wave 1 (wearout). For each carrier route, the sum of the degrees of membership to three clusters is equal to one, so we consider only *Membership to Cluster 1* and *Membership to Cluster 2* in our analysis.

[9] We use a symmetrized Kullback-Leibler divergence to measure similarity between the response timing distributions (Kullback and Leibler 1951).

**Figure 2. Average Response Rates by Day for the Three Clusters**




The figure reports average response rates to the fall pre-treatment mailings. The response rates are averaged across carrier routes weighted by the degrees of membership to different clusters. To protect confidentiality the response rates are indexed to 5% in Cluster 1 on day 1. The Wave 2 mailing date in the fall pre-treatment was approximately 37 days after the pre-treatment Wave 1 mailing date.

### Calculation of the Timing Variables

Recall that the purpose of the pre-treatment mailings is to measure whether the response rate increases or decreases between the two pre-treatment mailing waves (in each carrier route). Reliably measuring the slope of the response rate requires that there are sufficient responses to the fall pre-treatment mailing. Therefore, we initially only construct the three timing measures for the carrier routes that had at least ten responses to the fall pre-treatment mailings. These 394 carrier routes include 199,393 households and over 77% of the responses to the fall mailings. Notice that this filtering occurs prior to the spring treatment, and so it does not introduce a risk of selection bias due to confounds with the treatments themselves. We will also conduct extensive sensitivity analysis to investigate how the findings vary when we change this minimum response threshold.

Although we only construct timing measures for carrier routes with at least ten responses in fall, we retain all 796 carrier routes in our analysis of whether the timing measures explain variation in the spring response rates and whether they improve the performance of targeting models. We discuss this analysis next.

**Do the Timing Measures Explain the Variation in Spring Responses?**

To evaluate whether the timing of the pre-treatment fall responses helps to predict the spring treatment responses, we estimate the following OLS model:

$$Treatment_c - Control_c = \alpha + \boldsymbol{\tau}\ \boldsymbol{Timing_c} + \beta_1\ Fall\ Response_c + \beta_2\ Missing\ Timing_c + \boldsymbol{\beta}\ \textbf{Demographics}_c + \varepsilon_c \quad (1)$$

The unit of analysis in this model is a carrier route (*c*). The *Treatment* and *Control* variables measure the response rates in the respective treatment and control (no mail) conditions. We estimate this model separately for each of the four spring mailing treatments. Therefore, the dependent variable measures the lift in the response rate for that treatment compared to the control condition.

The ***Timing*** term refers to the timing variables. The coefficient for these variables is the coefficient of interest. In Model 1 we use the *Response Before Wave 2* as the timing variable. Larger values of this variable identify carrier routes in which a larger proportion of overall response occurred before the second mailing wave. These are carrier routes for which we would expect that the response to the free trial promotion is wearing out. Therefore, we anticipate a negative coefficient for the timing variable in this model. The coefficient $\tau$ of the timing variable has the following interpretation: other things equal, an increase by 100% in the proportion of responses before Wave 2 in fall will decrease the lift in response rate in spring by $|\tau|$ percentage units on average.

In Model 2 we use the *Average Response Date* as the timing variable. Larger values of this variable identify carrier routes in which a larger proportion of the overall fall response occurred later in fall. This is the opposite of the *Response Before Wave 2* and so we expect a positive coefficient on the timing variable in this model. The coefficient $\tau$ of the timing variable has the following interpretation: other things equal, an increase by one day in the *Average Response Date* in fall will increase the lift in response rate in spring by $\tau$ percentage units on average.

In Model 3 we use *Membership to Cluster 1* and *Membership to Cluster 2* as the timing variables.[10] The coefficients of these variables ($\tau$) have the following interpretation: other things equal, an increase by 100% will increase the lift in response rate in spring by $\tau$ percentage units on average.

[10] Recall that the sum of the 3 cluster memberships is 1, hence we only need to include two of the three membership variables in our model.

Because the *Membership to Cluster 3* variable is omitted from Equation 1, the *Membership to Cluster 1* coefficient measures the impact on the response rate if a carrier route looks more like Cluster 1 and less like Cluster 3. This is a shift towards the cluster associated with wearin and away from the cluster associated with wearout. This suggests we should see a positive coefficient for the *Membership to Cluster 1* variable.

In contrast, we do not have a clear prediction for the sign of the *Membership to Cluster 2* coefficient. This coefficient measures the impact on the response rate if a carrier route looks more like Cluster 2 and less like Cluster 3. Cluster 2 indicates a low response to the fall mailings, while Cluster 3 indicates wearout. Because both indications signal a lower response to the spring treatments, the impact on the spring treatment response is ambiguous.

We also include several additional variables as controls. The *Fall Response* measures the pre-treatment response rate in that carrier route. The **Demographics** term refers to the ten demographic control variables described in Section 3 (and in the Appendix). The *Missing Timing* is a binary dummy variable identifying the carrier routes for which fewer than ten responses were received to the fall pre-treatment mailings.[11] Recall that we only estimated the timing measures for carrier routes that had at least ten responses in the fall. The inclusion of the *Missing Timing* variable enables us to estimate Equation (1) using all 796 carrier routes. This means that the coefficients for the *Fall Response* and *Demographic* variables are identified using all 796 carrier routes.

We estimated Equation (1) separately for each of the four treatments. We also estimated a fifth model, where we average the response rate across all four spring mailing treatments (and subtract the response rate in the spring control condition from this average). We report the findings in Table 2, where to protect confidentiality we only report the coefficients of interest (the timing variables).[12]

The findings confirm that the timing of the responses to the fall mailings helps to explain variation in the responses to the spring treatment mailings. In Model 1 the coefficients for the *Response Before Wave 2* are all negative. In Model 2 the coefficients for the *Average Response Date* are all positive. Finally, in Model 3 the coefficients for the *Membership to Cluster 1* variable are also all positive. These findings are consistent with our interpretation that in carrier routes in which the promotion is still wearing in, the response rate will be higher in spring, but if the promotion is wearing out, the response rate will be lower in spring.

Although the evidence in this section is reassuring, it is merely preliminary. Our goal is to show that the timing of the responses in fall can help the firm target which carrier routes to mail to in spring. We address this goal in the next section by comparing two simple targeting models.

[11] The timing variables are set to zero for these carrier routes.
[12] The retailer prefers to reveal as little information as possible about the response to its promotions.

**Table 2. Describing Treatment Responses Using the Timing of Pre-Treatment Responses**

| | No Trial No | Trial No No | Trial Trial No | Trial Trial Trial | Average |
|---|---|---|---|---|---|
| **Model 1** | | | | | |
| Response Before Wave 2 | -2.775%* | -4.446%** | -4.560%** | -3.743%* | -3.881%** |
| | (1.355%) | (1.558%) | (1.597%) | (1.576%) | (1.115%) |
| **Model 2** | | | | | |
| Avg. Response Date | 0.028% | 0.035%† | 0.058%** | 0.045%* | 0.042%** |
| | (0.018%) | (0.021%) | (0.021%) | (0.021%) | (0.015%) |
| **Model 3** | | | | | |
| Membership to Cluster 1 | 2.715%** | 2.847%* | 2.570%* | 2.934%** | 2.766%** |
| | (0.960%) | (1.108%) | (1.136%) | (1.118%) | (0.792%) |
| Membership to Cluster 2 | -0.255% | 0.268% | 0.013% | 2.211% | 0.559% |
| | (1.290%) | (1.489%) | (1.527%) | (1.503%) | (0.106%) |

The table reports the coefficients of interest when estimating Equation (1). The dependent variable measures the average response rate for the respective treatment in that carrier route minus the average response rate for the control (no mail) condition in the same carrier route. The unit of analysis is a carrier route and the sample size is 796 carrier routes in all five models. Standard errors are in parentheses.

## 5. Targeting Models

In this section we investigate whether adding variables describing the timing of the fall pre-treatment responses can help improve the performance of a targeting model. The targeting model that we use is based upon the OLS model reported in the previous section. We emphasize that our goal is not to contribute to the sophistication of targeting models. Instead, the goal is to see whether the inclusion of the timing variables improves performance.

### Two Simple Targeting Models

We construct and validate two targeting models using a cross validation approach. In particular, for each treatment we use the following procedure:

1. Omit a single carrier route from the 796 carrier routes and use this as a holdout sample.
2. Using the remaining 795 carrier routes estimate the following two OLS models:

**Base Model**

$$Treatment_c - Control_c = \alpha + \beta_1\, Fall\ Response_c + \boldsymbol{\beta}\ \textbf{Demographics}_c + \varepsilon_c$$

**Timing Model**

$$Treatment_c - Control_c = \alpha + \boldsymbol{\tau}\ \boldsymbol{Timing_c} + \beta_1\, Fall\ Response_c + \beta_2\, Missing\ Timing_c + \boldsymbol{\beta}\ \textbf{Demographics}_c + \varepsilon_c$$

3. Using the parameters from the Base Model and Timing Model, predict the increase in response from mailing to households in the holdout carrier route. We label this increase *Predicted Base Lift* and *Predicted Timing Lift* (respectively).
4. Repeat steps 1 through 3 using each of the 796 carrier routes as a holdout sample.

While we start by using a single carrier route in each holdout sample, we will later also investigate increasing the holdout samples to include 5%, 10% and 20% of the 796 carrier routes. We note that increasing the size of the holdout sample reduces the size of the estimation samples, reducing the information available to estimate the Base and Timing Models. This also provides an indication of whether the three timing measures can (robustly) capture the necessary information even when fewer data is available.

To compare the performance of the Base and Timing Models we first consider the practical problem faced by the firm that provided data for this study. When making targeting decisions, this firm first sets a total number of households that it wants to mail to, and then decides which carrier routes to mail to in order to meet this target. Therefore, we will start by comparing how well the two models identify the best carrier routes to target given a fixed number of carrier routes to mail to. We will later allow the models themselves to choose how many carrier routes to mail to.

**Targeting a Fixed Number of Carrier Routes**

To choose a fixed number of carrier routes to mail to we use the predicted values from the two models, *Predicted Base Lift* and *Predicted Timing Lift*. For each treatment and timing measure we use the following procedure:

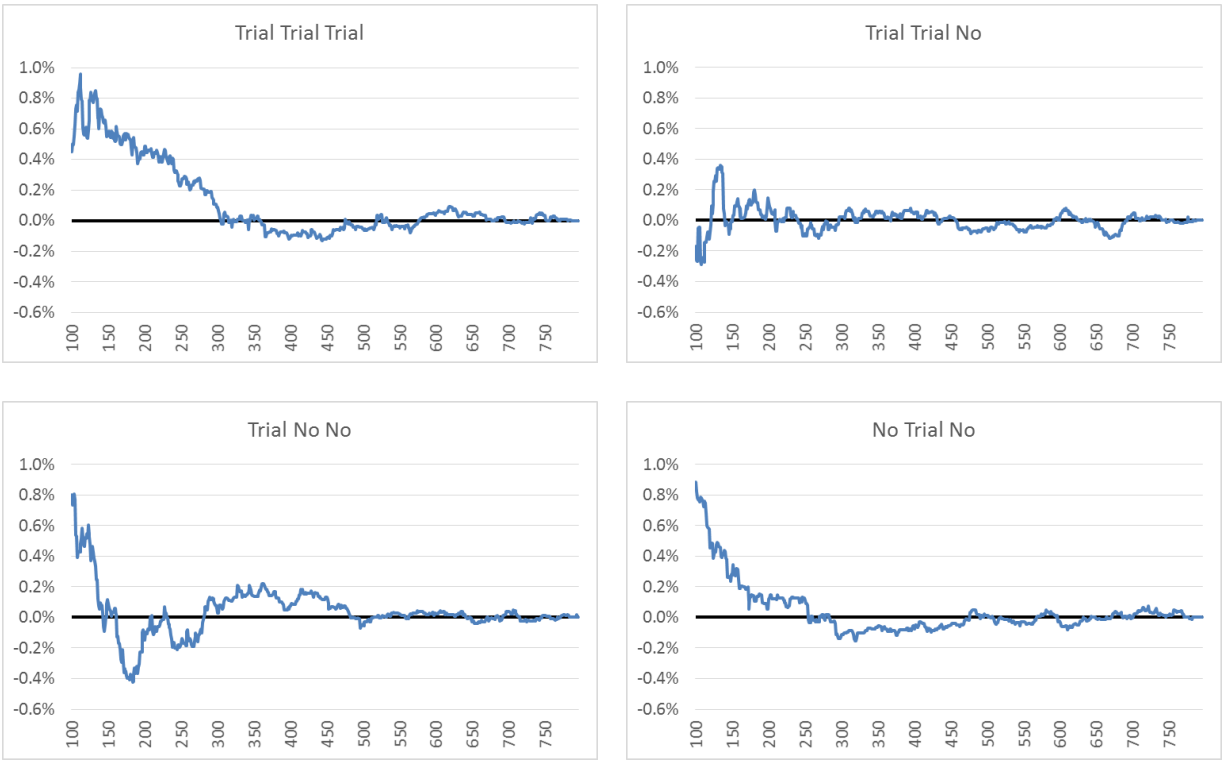
1. Rank the carrier routes in decreasing order of *Predicted Base Lift*, and capture the rank order as *Predicted Base Rank*. We similarly use the *Predicted Timing Lift* to construct the *Predicted Timing Rank*. These ranks represent the priority in which each targeting model would mail to the 796 carrier routes.[13]

[13] Notice that because this ranking is constructed separately for each treatment, the mailing costs and expected profit per response are not relevant because they are the same for each carrier route.

2. For each rank, 1 through 796, construct the *Actual Base Lift* and *Actual Timing Lift* equal to the difference in the actual response rates in the Treatment and Control conditions in the carrier route assigned to that rank by the *Predicted Base Rank* and *Predicted Timing Rank*.
3. For each rank, calculate the difference between *Actual Timing Lift* and *Actual Base Lift*. We label this difference *Increase in Lift*.

We use the *Increase in Lift* to evaluate the performance of the models. As we describe above, this is the actual lift in the response rate produced by the Timing Model minus the actual lift produced by the Base Model, by rank order. Positive values indicate that the Timing Model chose carrier routes that yielded a greater lift in the response rate than the Base Model. In Figures 3a through 3d we report the average *Increase in Lift* when varying the total number of carrier routes to mail. The y-axis identifies the average *Increase in Lift*, while the x-axis identifies the total number of carrier routes mailed (varying from a minimum of 100 to all 796).

**Figures 3a-3d: Timing Model vs. Base Model**
***Increase in Lift* When Using the *Membership to Clusters* Timing Model**

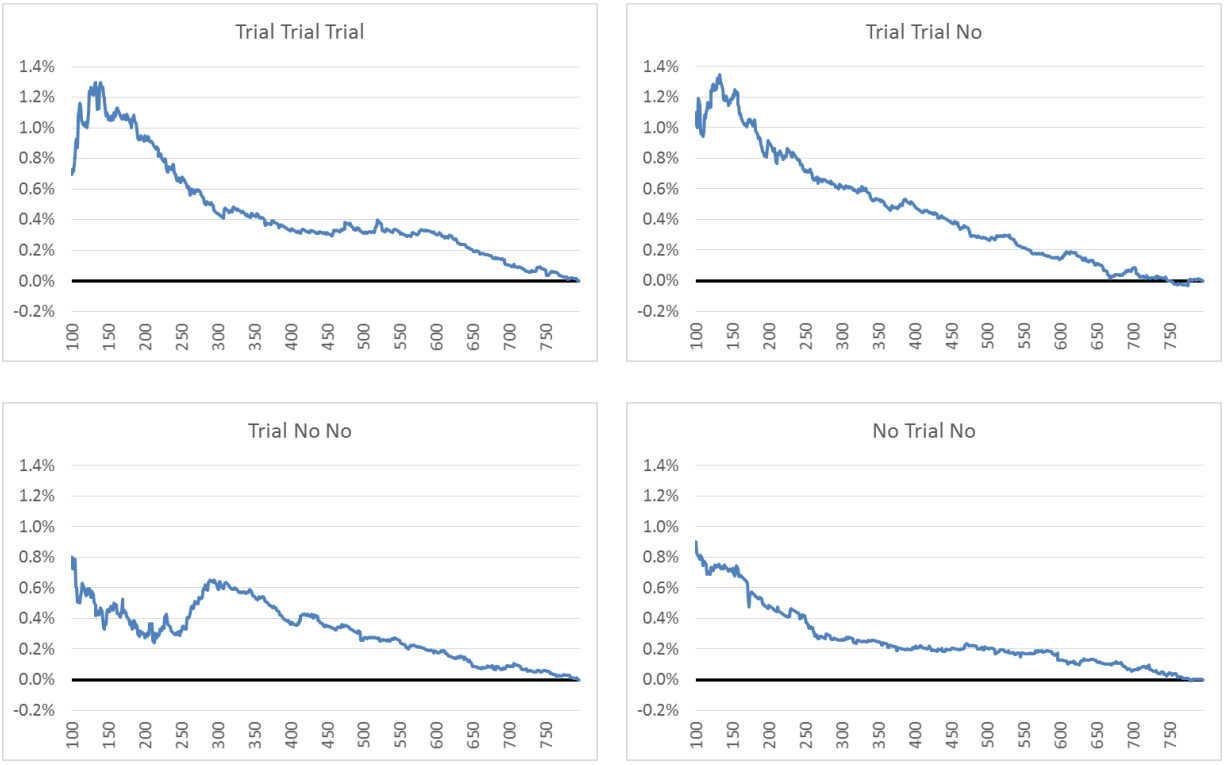



These figures report the difference in the lift in the response rate when using the Timing Model compared to the Base Model (*Increase in Lift*). The lift in the response rate for a carrier route is the increase in the actual response rate in the Treatment condition compared to the Control condition for that carrier route. The y-axis measures the average *Increase in Lift* for the carrier routes ranked highest by each model. Positive values indicate that the lift in the response rate was higher when using the Timing Model compared to the Base Model. The x-axis identifies the total number of top-ranked carrier routes used to calculate the *Increase in Lift*. The Timing Model used to construct these figures uses the *Membership to Clusters* timing variables.

We also compare the Timing Model with a Naïve Benchmark. The Naïve Benchmark is the average lift in response rate across all 796 carrier routes. The performance of this benchmark is equivalent to randomly selecting a fixed number of carrier routes to mail to. These findings are reported in Figures 4a through 4d.

In all of these figures we focus on the Timing Model estimated using the *Membership to Clusters*. We will later compare the performance of all three timing variables.

**Figures 4a-4d: Timing Model vs. Naïve Benchmark**
***Increase in Lift* When Using the *Membership to Clusters* Timing Model**




These figures report the difference in the lift in the response rate when using the Timing Model compared to the Naïve benchmark. The lift in the response rate for a carrier route is the increase in the actual response rate in the Treatment condition compared to the Control condition for that carrier route. The Naïve Benchmark is the average increase in response rates in the Treatment Condition compared to the Control (across all 796 carrier routes). The y-axis measures the average increase in response rates in the Timing Model minus the Naïve Benchmark for the carrier routes ranked highest by the Timing Model. Positive values indicate that the lift in the response rate was higher in the Timing Model than the Naïve Benchmark. The x-axis identifies the total number of top-ranked carrier routes used to calculate the lift in the response rate when using the Timing Model. The Timing Model used to construct these figures uses the *Membership to Clusters* timing variables.

The figures reveal that when selecting the top 100 to 150 carrier routes to mail to, the Timing Model performs noticeably better than the Base Model. Within this range the average *Increase in Lift* is positive for three of the four treatments.

The exception is the Trial Trial No treatment. For this treatment, the *Membership to Clusters* Timing Model does not generate any increase in the response rate compared to the Base Model. This finding is somewhat consistent with our earlier OLS results. Recall that in Table 2, the *Membership to Clusters* timing variables were smaller in magnitude (and less statistically significant) in this treatment than in the other treatments.

As the number of target carrier routes increases, the improvement in performance disappears. This is not surprising. The treatments did not generate any responses in many carrier routes. As a result, when targeting a large number of carrier routes, the models necessarily choose many carrier routes for which the treatments yield no responses. This dilutes the difference in the performance of the models. At the limit, when choosing all 796 carrier routes, the performance of the two targeting models is identical (by construction).

The comparisons in Figures 4a – 4d demonstrate that the Timing Model also consistently outperforms the Naïve Benchmark, both when varying the treatments, and when varying the fixed number of carrier routes to target.

In Figures 3a – 3d we calculated the difference in the performance of the Timing and Base models using all of the carrier routes selected by the two models. However, we note that in many cases the two models chose to target the same carrier routes. For example, in the No Trial No treatment, the first 100 carrier routes include 64 carrier routes chosen by both models. The other 36 carrier routes do not overlap. In Table 3 we compare the outcomes for these different groups of carrier routes.

Although the total difference using all 100 carrier routes is not statistically significant, when we focus on the 36 carrier routes for which the two models make different mailing decisions, then the difference is significant. We can conduct the same analysis for all three Timing Models and when varying the total number of carrier routes to target. In Table 4 we report the findings when targeting 100, 150 or 200 carrier routes. In the Appendix we also report the findings when comparing the Timing Model with the Naïve Benchmark.

For all four treatments the Timing Model consistently yields a larger lift in the response rate when selecting the top 100 carrier routes. This pattern holds irrespective of which treatment and which timing variable we use (the only exception is the Trial Trial No treatment in the *Membership to Clusters* model).

**Table 3: Targeting 100 Carrier Routes Using the *No Trial No* Treatment**

| | Timing Model | Base Model | Difference (*Increase in Lift*) |
|---|---|---|---|
| All 100 carrier routes | 1.180% (0.486%) | 0.294% (0.418%) | 0.887% (0.588%) |
| 64 carrier routes chosen by both models | 0.594% (0.538%) | 0.594% (0.538%) | 0.0% |
| 36 different carrier routes | 2.224% (0.937%) | -0.240% (0.662%) | 2.463%* (1.148%) |

This table compares the lift in the response rate when using the Timing and Base Models to select 100 of the 796 carrier routes to target. The lift in response in the first row is calculated as the average increase in the response rate in the No Trial No Treatment condition compared to the Control condition for the 100 carrier routes ranked highest by each model. The Timing Model uses the *Membership to Clusters* timing variables. Standard errors are in parentheses.

However, this effect only survives when focusing on the 100 highest ranked carrier routes. When we extend the focus to include the first 150 or the first 200 ranked carrier routes, we no longer see a larger lift in response rates in the Timing Model.

The significance of the effect also varies across the three timing measures. The *Response Before Wave 2* and *Membership to Clusters* Timing Models both yield significant improvements over the Base Model when targeting 100 carrier routes. In contrast, the *Average Response Date* Timing Model does not yield a significant improvement either for any of the individual treatments or when aggregating across the four treatments.

In additional analysis we also focused on the last carrier routes that each model would choose to mail to. Reassuringly, the carrier routes that the Timing Model ranked last tended to have lower lifts in response than the carrier routes that the Base Model ranked last. However, these differences were not statistically significant. Many of these low ranked carrier routes had no responses in either the Treatment or Control conditions, and so the modal lift in response rates was zero.

These findings confirm that we can use the timing of past responses by responders to help target non-responders. We continue our analysis by further investigating the robustness of the findings, and then calculate the impact on firm profits.

**Table 4: Targeting a Fixed Number of Carrier Routes**
**Difference Between Timing and Base Models**

| | | Top 100 | Top 150 | Top 200 |
|---|---|---|---|---|
| Response Before Wave 2 | No Trial No | 2.406%<br>(1.552%) | 0.363%<br>(1.666%) | 0.341%<br>(1.334%) |
| | Trial No No | 2.147%$^{\dagger}$<br>(1.238%) | 1.645%<br>(1.114%) | 0.657%<br>(0.927%) |
| | Trial Trial No | 0.654%<br>(1.506%) | 0.081%<br>(0.948%) | 0.217%<br>(0.707%) |
| | Trial Trial Trial | 1.602%<br>(1.174%) | 0.004%<br>(0.766%) | 0.278%<br>(0.751%) |
| | Average | 1.679%$^{**}$<br>(0.669%) | 0.523%<br>(0.529%) | 0.365%<br>(0.440%) |
| Average Response Date | No Trial No | 1.072%<br>(1.439%) | -1.088%<br>(1.811%) | -0.157%<br>(1.480%) |
| | Trial No No | 0.253%<br>(1.330%) | -1.222%<br>(1.648%) | -1.216%<br>(1.237%) |
| | Trial Trial No | 0.198%<br>(1.292%) | -2.554%$^{*}$<br>(1.282%) | -0.306%<br>(0.852%) |
| | Trial Trial Trial | 0.172%<br>(1.234%) | 0.335%<br>(0.936%) | 0.343%<br>(0.833%) |
| | Average | 0.384%<br>(0.664%) | -0.900%<br>(0.646%) | -0.181%<br>(0.508%) |
| Membership to Clusters | No Trial No | 2.463%$^{*}$<br>(1.148%) | 0.878%<br>(0.983%) | 0.556%<br>(0.957%) |
| | Trial No No | 2.226%$^{\dagger}$<br>(1.234%) | 0.302%<br>(1.014%) | -0.536%<br>(1.011%) |
| | Trial Trial No | -0.495%<br>(1.507%) | -0.040%<br>(0.872%) | 0.609%<br>(0.835%) |
| | Trial Trial Trial | 1.023%<br>(1.084%) | 1.323%$^{\dagger}$<br>(0.756%) | 1.269%$^{\dagger}$<br>(0.695%) |
| | Average | 1.313%$^{*}$<br>(0.618%) | 0.717%<br>(0.452%) | 0.541%<br>(0.436%) |

This table reports the difference in the lift in the response rate when using the Timing Model compared to the Base Model (*Increase in Lift*). Positive values indicate that the lift in the response rate was higher when using the Timing Model. The lift in the response rate is calculated as the average increase in the response rate in the Treatment condition compared to the Control condition for the carrier routes ranked highest by each model. We omit carrier routes selected by both models. Standard errors are in parentheses.

**Robustness**

We conducted two sets of robustness checks. They both focus on the performance of the models when mailing to a total of 100 carrier routes. Detailed findings are reported in the Appendix.

First, we re-estimated the models when constructing the timing measures using different minimum thresholds of fall responses. The findings are robust to estimating the timing variables using fewer fall responses. Even when using a threshold as low as 5 pre-treatment responses, we continue to see that the Timing Model outperforms the Base Model. This replication is reassuring, and suggests that the timing of pre-treatment responses can help improve targeting even if there are relatively few past responses available to evaluate whether a promotion is wearing in or wearing out.

On the other hand, the improvement in the performance of the targeting model is less apparent when using a minimum threshold of 12 or 15 responses in fall. This reduction in performance is not surprising as the timing variables are only estimated using a relatively small proportion of the carrier routes.

Our second robustness check repeated the cross validation procedure using holdout samples of different sizes. Recall that in the results in Table 4 we use a holdout sample of a single carrier route (rotating across all 796 carrier routes). We replicate the analysis when targeting 100 carrier routes using holdout samples of 5%, 10% or 20% of the 796 carrier routes (again rotating across all of the carrier routes). Although the pattern of findings is unchanged, the statistical significance of the results is diminished when using larger holdout samples. With larger holdout samples there are smaller estimation samples to estimate the various models (for example, with a 20% holdout sample the estimation sample is just 80% of the 796 carrier routes).

Our final analysis using a fixed total mailing target evaluates the impact on firm profits. We discuss these findings next.

**Profit Improvement**

We again focus on the performance of the models when mailing to a total of 100 carrier routes. Profits are calculated using actual transactions in the one year after the date of the first spring mailing date. The profit calculations use a flat percentage profit margin (provided by the firm), the actual mailing costs, and the actual transactions in each treatment in each carrier route over this one-year period. To protect confidentiality the profits are multiplied by a common random number.

The findings are reported in Table 5, where we report the difference in profits earned from the Timing and Base Models. We omit carrier routes that were chosen by both targeting models.

The findings reveal that the Timing Models consistently yielded higher profits than the Base Model. However, there was a lot of variance in the amount that each new customer spent in the

firm's stores over the one-year measurement period, and so the average profit measures have large variances. As a consequence, the profit improvements are only statistically significant for some of the treatments.

**Table 5. Profit Improvement over Base Model When Targeting 100 Carrier Routes**

| | Response Before Wave 2 | Average Response Date | Membership to Clusters |
|---|---|---|---|
| No Trial No | $9.51 ($7.88) | $5.23 ($8.14) | $0.84 ($7.03) |
| Trial No No | $15.58† ($7.97) | $2.55 ($8.73) | $13.96* ($5.81) |
| Trial Trial No | $7.06 ($9.29) | -$0.64 ($8.84) | $0.19 ($7.27) |
| Trial Trial Trial | $10.12 ($9.14) | $3.16 ($10.52) | -$0.57 ($8.83) |
| Average | $10.64* ($4.38) | $2.47 ($4.84) | $3.43 ($3.76) |

The table reports the difference in average profit in the Timing and Base Models when using the Timing and Base Models to select 100 of the 796 carrier routes to target. The profits of a treatment in a carrier route are calculated using the actual profit earned using the corresponding treatment in that carrier route. To protect confidentiality the profits are multiplied by a common random number. We omit carrier routes that were selected by both the Timing and the Base Model. The unit of analysis is a carrier route. Standard errors are in parentheses.

While our findings confirm that the timing of past responses has the potential to improve targeting decisions when mailing to a pre-determined number of carrier routes, this is not always the problem that firms face. A firm may also want to allow the models to choose how many carrier routes to mail to. We compare the performance of the Timing and Base Models on this problem next.

**Choosing the Optimal Number of Carrier Routes to Target**

In this analysis we use the same Timing and Base Models, but instead of ranking the carrier routes in order of priority, we now make a separate mailing decision for each carrier route. In particular, we use the Timing Model to assign mailing decisions for each carrier route (*c*) and treatment (*e*) using the following decision rule:

> "Mail if *Predicted Timing Lift*$_{ce}$**Average Profit per Response*$_e$ is larger than the mailing cost for that treatment (*e*), and do not mail otherwise."

The *Average Profit per Response*$_e$ is calculated using all of the households that responded to that treatment. We separately use the *Predicted Base Lift*$_{ce}$ to make mailing decisions for the Base Model.

We also identify the "correct" mailing decisions using the following decision rule:

> "Mail if *Actual Treatment Profit*$_{ce}$ – *Actual Control Profit*$_c$ is larger than the mailing cost for that treatment, and do not mail otherwise."

In Table 6 we compare how well the Timing and Base Models identify the "correct" mailing decisions. We focus on the carrier routes in which the Timing and Base Models recommended different mailing decisions and report the number of times these models chose the correct mailing decision. We also report the results of a binomial test to compare whether the Timing Model chooses the correct action more often than the Base Model. In particular, we report the p-value from a two-tail test evaluating whether one model is correct more or less frequently than the other model.[14] The findings again confirm that inclusion of the timing variables improves the targeting model. The percentage of correct mailing decisions is higher in the Timing Model than in the Base Model. The findings are consistent across the three timing measures, with the *Response Before Wave 2* Timing Model yielding the largest performance improvements.

The one treatment in which the Timing Models do not improve upon performance is the No Trial No treatment. Recall that this is the treatment that yielded the smallest lift in responses. For this treatment the difference in profit between the Treatment and Control conditions is relatively small. As a result, the outcome measure is relatively noisy because small variations in spending by the new customers can easily change whether a decision to mail was "correct". Consistently with this argument, the Timing Models provided the largest improvement over the Base Model in the Trial Trial Trial condition: this is the condition in which there was the largest difference in response between the Treatment and Control models.

We conducted the same robustness checks on this analysis as we implemented when choosing a fixed number of carrier routes to mail to. The findings of both robustness checks are reported in the Appendix. They reveal that when allowing the models to choose how many (and which) carrier routes to mail to, the three timing models are relatively robust to the number of pre-treatment responses. The findings when varying the minimum threshold fall responses reveal a similar pattern to the findings in Table 6.

[14] This is an exact test of the statistical significance of deviations from an expected distribution of observations where the outcome can fall into two categories. For example, it could be used to evaluate whether a coin is "fair" by testing whether the heads outcome occurs significantly more frequently than the tails outcome.

The findings are also reassuringly robust to varying the size of the holdout sample. Even when using a holdout sample as large as 20%, we see that all three Timing Models yield more accurate mailing decisions than the Base Model.

**Table 6. Comparison of Timing and Base Models**

| Timing Measure | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| Response Before Wave 2 | No Trial No | 29 | 36 | 0.3853 |
| | Trial No No | 34 | 19 | 0.0394* |
| | Trial Trial No | 40 | 29 | 0.1854 |
| | Trial Trial Trial | 68 | 35 | 0.0011** |
| | Total | 171 | 119 | 0.0023** |
| Average Response Date | No Trial No | 16 | 26 | 0.1228 |
| | Trial No No | 19 | 12 | 0.2087 |
| | Trial Trial No | 38 | 33 | 0.9055 |
| | Trial Trial Trial | 63 | 38 | 0.0129* |
| | Total | 136 | 109 | 0.0845† |
| Membership to Clusters | No Trial No | 42 | 51 | 0.3507 |
| | Trial No No | 21 | 16 | 0.4111 |
| | Trial Trial No | 30 | 28 | 0.7928 |
| | Trial Trial Trial | 82 | 53 | 0.0126* |
| | Total | 175 | 148 | 0.1330 |

The table reports the number of comparisons in which the Timing Model and Base Model make the correct targeting recommendation. The unit of observation is a carrier route. The correct decision is identified using the actual profits in the Control and corresponding Treatment condition. We restrict attention to carrier routes in which the two models make different recommendations. The table reports the p-value from a two-tail test evaluating whether one model is correct more or less frequently than the other model.

**Summary**

We have compared three Timing Models constructed using different versions of the timing variables. The models were compared against a Base Model that was identical to the Timing Model, with the exception that the timing variables were omitted from the Base model. We compared the models on their ability to target the most attractive carrier routes either when the goal is to mail a fixed number of carrier routes, or when the models choose how many carrier routes to target.

The findings reveal consistent evidence that the inclusion of the timing variables in the targeting model improved performance. This is particularly true for the *Response Before Wave 2* and the *Membership to Clusters* models. While the *Membership to Clusters* model is more complicated to construct, aggregating the responses in each cluster also provides an intuitive (and visual) segmentation of the response curves. The performance of the *Response Before Wave 2* model has an important practical implication. For firms seeking to use the timing of past responses in their targeting models, it may be sufficient to use a relatively simple timing measure.

Finally, a firm may be concerned not just with whether to mail, but also how frequently and when to mail. Choosing the timing and frequency of mailings can also be thought of as a targeting problem in our setup, with the firm choosing which of the four treatments to target each carrier route with. In our study, the predicted profitability of the Trial No No treatment dominates the other three treatments in almost every carrier route.[15] As a result, the frequency and timing decisions become a simple comparison of Trial No No and the (no mail) control, which we have already evaluated.

## 6. Conclusions

A fundamental challenge when prospecting for new customers is that the customers who responded to past promotions are no longer in the mailing pool. As a result, the mailing pool becomes increasingly diluted and there is no variation in the behavior of the remaining households to help decide whom to target with future promotions.

In this paper we investigate whether we can use past responders' behavior to help target non-responders. Our findings confirm that the timing of responses by past responders can improve targeting of non-responders in future mailing decisions. We observe an improvement both in the number of decisions that are correct, and in the average profits earned.

We have interpreted this effect as evidence of wearin or wearout of a promotion. If the response rate increases over multiple waves of a promotion, this signals that the promotion in a region is still wearing in, which suggests a larger response to future mailings. In contrast, if the response rate decreases across multiple waves of a promotion, this may indicate that the promotion is wearing out in that region, and so future response rates will also be lower. Although this interpretation is consistent with our findings, we acknowledge that we have no direct evidence that this is due to wearin or wearout of the promotion. In particular, we have not tested the psychological mechanisms that contribute to wearin and wearout of a promotion. Testing these

[15] Recall from our earlier discussion that the Trial No No treatment yielded a higher response than the No Trial No treatment, for the same mailing cost. The Trial No No treatment also provides a similar lift in responses to the two treatments that mailed more frequently (Trial Trial No and Trial Trial Trial), while incurring lower mailing costs.

mechanisms is beyond the capabilities of our experimental data, and so beyond the scope of this study.

While we have used a direct mail setting in this study, the findings may extend to other settings, including retargeting of digital advertising. Retargeting describes the widely used practice of using previous online browsing history to select which advertising content to display. For example, if a customer inspects a sweatshirt on a Gap website but does not purchase it, the customer may be shown digital advertisements for the same sweatshirt (or other Gap products) when visiting subsequent websites. The focus on customers who did not initially purchase is similar to this study's retailer's focus on prospects that did not previously respond. This characteristic of retargeting has led to concerns that retargeting is less profitable than other types of digital advertising (see for example Lambrecht and Tucker 2009). Like the retailer in our study, digital advertisers must decide which customers to stop retargeting because they are unlikely to ever respond.

## 7. References


Allenby, G. M., and Rossi, P. E. (1998), Marketing models of consumer heterogeneity, *Journal of Econometrics*, 89(1), 57-78.

Anderson, Eric T. and Duncan I. Simester (2001), Research note: Price discrimination as a signal: Why an offer to spread payments may hurt demand, *Marketing Science,* 20(3), 315-327.

Ansari, A., and Mela, C. F. (2003), E-customization, *Journal of Marketing Research*, 40(2), 131-145.

Ascarza, Eva (2017), Retention futility: Targeting high risk customers might be ineffective, forthcoming at the *Journal of Marketing Research*.

Bass, F. M., Bruce, N., Majumdar, S., and Murthi, B. P. S. (2007), Wearout effects of different advertising themes: A dynamic Bayesian model of the advertising-sales relationship, *Marketing Science*, *26*(2), 179-195.

Bezdek, James C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms,* Springer US.

Brynjolfsson, E., Y. (J.) Hu and D. Simester, Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales, *Management Science*, 57(8), 1373-1386.

DMA (2006), *Statistical Fact Book*, Direct Marketing Association, New York NY.

DMA (2015), *Response Rate Report*, Direct Marketing Association, New York NY.

Dubé, J.-P., Z. Fang, N. M. Fong, and X. Luo (2017), Competitive price targeting with smart-phone coupons, forthcoming at *Marketing Science*.

Dubé, Jean-Pierre, and Sanjog Misra (2017), Scalable price targeting, working paper.

Dunn, J. C. (1973), A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics*, 3 (3): 32-57.

Gini C. (1955), *Variabilite Mutabilita*. In: Pizetti, E, et. al. (eds). Rome, IT: Libreria Eredi Virgilio Veschi.

Goel, Sharad, and Daniel G. Goldstein (2013), Predicting individual behavior with social networks, *Marketing Science,* 33(1), 82-93.

Kullback, S., and Leibler, R. A. (1951), On information and sufficiency, *The Annals of Mathematical Statistics*, 22(1), 79-86.

Li, J. Q., Rusmevichientong, P., Simester, D., Tsitsiklis, J. N., and Zoumpoulis, S. I. (2015), The value of field experiments, *Management Science*, 61(7), 1722-1740.

Lorenz, M. O. (1905), Methods of measuring the concentration of wealth, *Publications of the American Statistical Association*, 9 (70), 209-219.

Luo, X., Andrews, M., Fang, Z., and Phang, C. W. (2013), Mobile targeting. *Management Science*, 60(7), 1738-1756.

Naik, P. A., M. K. Mantrala, and A. G. Sawyer (1998), Planning media schedules in the presence of dynamic advertising quality, *Marketing Science*, 17(3) 214-235.

Pechmann, Cornelia and David W. Stewart (1988), Advertising repetition: A critical review of wearin and wearout, *Current Issues and Research in Advertising*, 11(2), 285-330.

Ray, M. L. and A. G. Sawyer (1971), Behavioral measurement for marketing models: Estimating the effects of advertising repetition for media planning, *Management Science*, 18(4, Part 2), 73-89.

Rossi, P. E., R. McCulloch, and G. M. Allenby (1996), The value of purchase history data in target marketing, *Marketing Science*, 15(4), 321-340.

Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2017), Customer acquisition via display advertising using multi-armed bandit experiments, *Marketing Science*, Articles in Advance.

Simester, D. (2017), Field experiments in marketing, *Handbook of Economic Field Experiments*, *1*, 465-497.

Simester, D. I., Sun, P., and Tsitsiklis, J. N. (2006), Dynamic catalog mailing policies, *Management Science*, *52*(5), 683-696.

Toubia, O., Goldenberg, J., and Garcia, R. (2014), Improving penetration forecasts using social interactions data, *Management Science*, *60*(12), 3049-3066.

Zhang, J., O. Netzer, and A. Ansari (2014), Dynamic targeted pricing in B2B relationships, *Marketing Science*, 33(3), 317-337.

Zhang, J. and Michel Wedel (*2009*), The effectiveness of customized promotions in online and offline Stores, *Journal of Marketing Research*, 46(2), 190-206.

# Using Past Responders to Target Non-Responders

## Appendix

September 2017

Theodoros Evgeniou
theodoros.evgeniou@insead.edu

Duncan Simester
simester@mit.edu

Artem Timoshenko
atimoshe@mit.edu

Spyros I. Zoumpoulis
spyros.zoumpoulis@insead.edu

## Appendix

### Definition of Descriptive Variables

| Variable | Definition | Level of Aggregation of Raw Data |
|---|---|---|
| Age | Age of head of household | Carrier Route |
| Home Value | Estimated home value | Carrier Route |
| Income | Estimated household income | Carrier Route |
| Single Family | A binary flag indicating whether the home is a single family home | Carrier Route |
| Multi-family | A binary flag indicating whether the home is a multi-family home | Carrier Route |
| Distance | Distance to nearest store for this retailer | Zip |
| Comp. Distance | Distance to nearest competitors' store | Zip |
| F Flag | Binary flag indicating whether the retailer considers the zip code "far" from its closest store | Zip |
| M Flag | Binary flag indicating whether the retailer considers the zip code a "medium" distance from its closest store | Zip |
| Nbr Households | Number of households in the carrier route | Carrier Route |

**Descriptive Variables Summary Statistics**

| Variable | Mean | Standard Error |
|---|---|---|
| Age | 56.71 | 0.23 |
| Home Value | 266,274 | 6,741 |
| Income | 82,981 | 2,234 |
| Single Family | 0.7206 | 0.0112 |
| Multi-family | 0.2750 | 0.0112 |
| Distance | 8.55 | 0.26 |
| Comp. Distance | 8.20 | 0.30 |
| F Flag | 0.49 | 0.02 |
| M Flag | 0.34 | 0.02 |
| Number Households | 572.32 | 7.42 |

The table reports summary statistics for the ten descriptive variables. The unit of analysis is a carrier route. The sample size for all ten variables is 796.

**Treatments Omitted From the Analysis**

The spring mailing experiment included a series of additional treatments designed to measure the impact of a second type of promotion, and of combining different sequences of the two promotions. In total there were 18 different treatment conditions plus the no mail control condition. In this study we focus on data from just four of the treatment conditions and the control condition.

We originally intended to replicate our analysis using the second type of promotion. The four treatments that we use in the analysis for the free trial promotion were also implemented for the second type of promotion. Moreover, approximately 400,000 of the households received fall pre-treatment mailings containing the second type of promotion. This mimicked the experimental design of the free trial promotion.

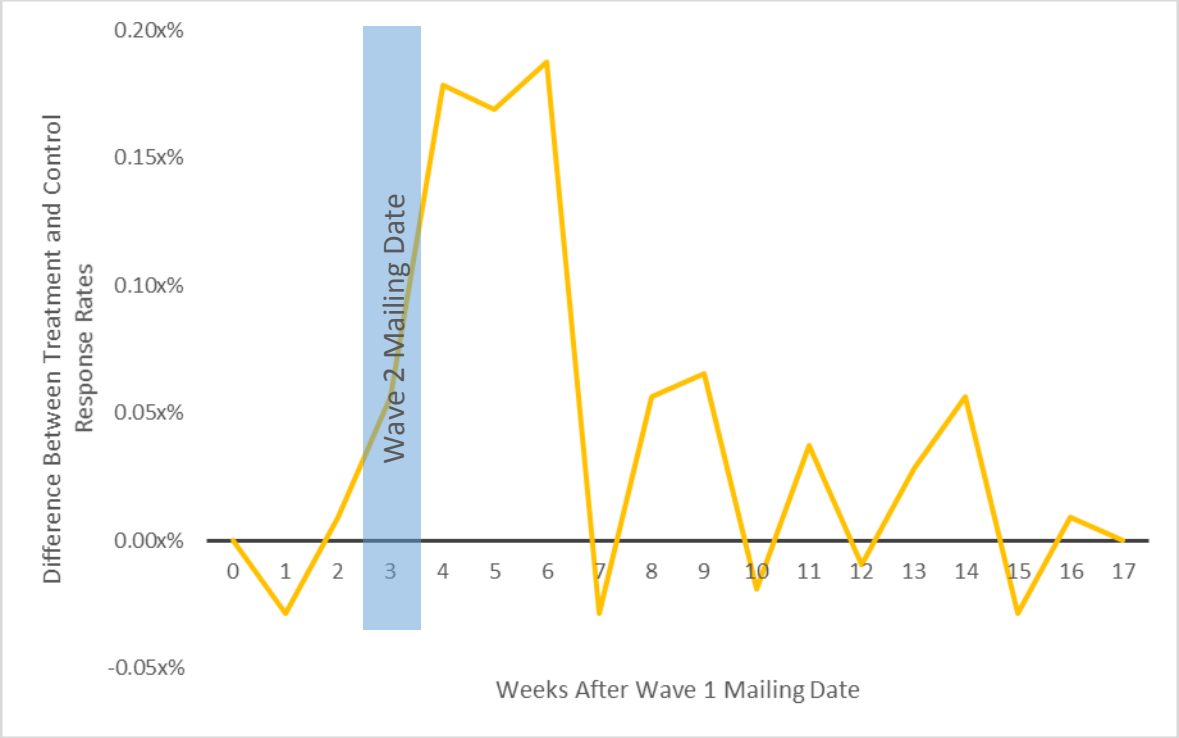
Unfortunately too few households responded in fall for us to reliably measure the timing of the response to this second promotion in fall. As a result we dropped the second type of promotion from our analysis before attempting to estimate a targeting model with this data. Fortunately the firm was able to use the results from these treatment conditions to make other decisions.

## Increase in Response Rates Attributable to the Spring Treatments

| | Lift in Response | Standard Error |
|---|---|---|
| No Trial No | 0.41% | 0.19% |
| Trial No No | 1.36% | 0.21% |
| Trial Trial No | 1.49% | 0.22% |
| Trial Trial Trial | 1.51% | 0.22% |

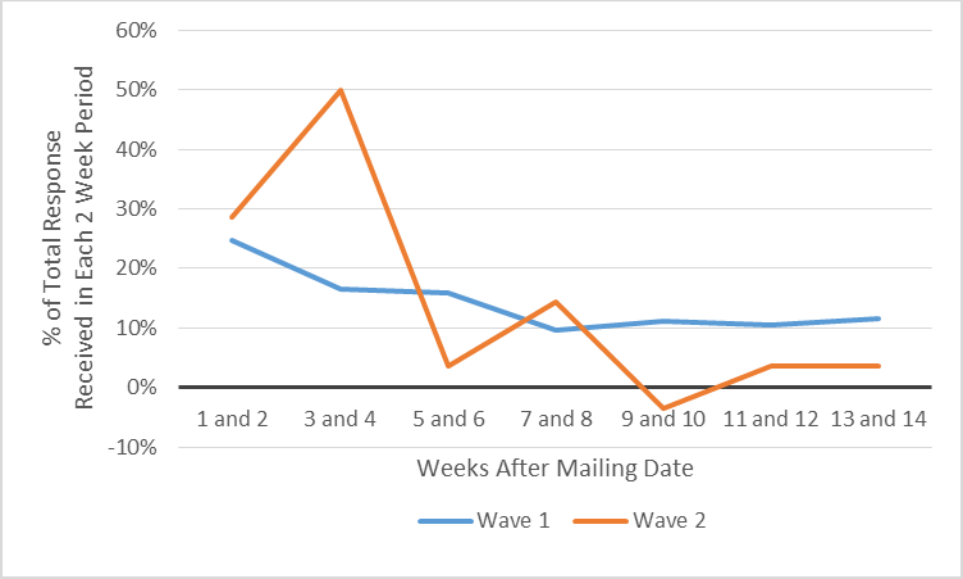
The table reports the difference between the response rate in each spring treatment condition and the response rate in the spring control condition. The response rate is calculated as the number of households that responded in that condition, divided by the total number of households in that condition. To preserve confidentiality the scale of the response rates is disguised (by multiplying by a random number). The unit of analysis is a carrier route and the sample size for each treatment is 796.

**Response Rate by Week in the No Trial No Condition Compared to the Control**




The figure reports the difference between the response rate in the No Trial No treatment condition and the response rate in the spring control condition. The response rate is calculated as the number of households that responded that week, divided by the total number of households in that condition. To preserve confidentiality the scale of the response rates is disguised (by multiplying by a random number). The line at the x-axis indicates zero difference in the response rates in the No Trial No and control condition. Sample sizes are reported in Table 1.

**Response Curve for Spring Wave 1 and Wave 2 Mailing Waves**




The figure reports the percentage of the total response received to each mailing wave in each two-week period. The response to Wave 1 is calculated using the response to the Trial No No spring treatment. The response to Wave 2 is calculated using the difference in the response to the Trial Trial No and Trial No No spring treatments.

**Targeting a Fixed Number of Carrier Routes**
**Difference Between Timing Model and Naïve Benchmark**

| | Top 100 | Top 150 | Top 200 |
|---|---|---|---|
| No Trial No | 0.900%$^{\dagger}$ | 0.705%$^{\dagger}$ | 0.485% |
| | (0.502%) | (0.362%) | (0.321%) |
| Trial No No | 0.801%$^{\dagger}$ | 0.452% | 0.273% |
| | (0.466%) | (0.404%) | (0.352%) |
| Trial Trial No | 1.096%$^{\dagger}$ | 1.187%$^{*}$ | 0.907%$^{*}$ |
| | (0.601%) | (0.498%) | (0.416%) |
| Trial Trial Trial | 0.692% | 1.052%$^{*}$ | 0.947%$^{*}$ |
| | (0.473%) | (0.461%) | (0.385%) |
| Average | 0.872%$^{**}$ | 0.849%$^{**}$ | 0.653%$^{**}$ |
| | (0.256%) | (0.218%) | (0.186%) |

This table reports the difference in the lift in the response rate when using the Timing Model compared to the Naïve Benchmark. Positive values indicate that the lift in the response rate was higher when using the Timing Model. The lift in the response rate is calculated as the average increase in the response rate in the Treatment condition compared to the Control condition for the carrier routes ranked highest by the Timing Model. Standard errors are in parentheses. The Timing Model is constructed using the *Membership to Clusters* timing variables.

**Varying the Minimum Fall Response Threshold**
**Targeting 100 Carrier Routes: Difference Between Timing and Base Models**

| | | 5 or More | 8 or More | 10 or More | 12 or More | 15 or More |
|---|---|---|---|---|---|---|
| Response Before Wave 2 | No Trial No | 2.557%* | 1.017% | 2.406% | 0.406% | -1.105% |
| | | (1.267%) | (1.250%) | (1.552%) | (0.914%) | (0.859%) |
| | Trial No No | 0.953% | 1.351% | 2.147%† | 1.008% | -0.911% |
| | | (1.913%) | (1.659%) | (1.238%) | (1.520%) | (1.774%) |
| | Trial Trial No | 1.884% | 1.203% | 0.654% | -0.029% | 1.420% |
| | | (3.112%) | (1.461%) | (1.506%) | (2.126%) | (1.164%) |
| | Trial Trial Trial | -7.177% | 1.319% | 1.602% | 0.399% | 0.221% |
| | | (4.629%) | (1.654%) | (1.174%) | (1.384%) | (1.278%) |
| | Average | 1.177% | 1.232% | 1.679%** | 0.408% | 0.181% |
| | | (1.066%) | (0.762%) | (0.669%) | (0.906%) | (0.714%) |
| Average Response Date | No Trial No | 2.548%* | 1.316% | 1.072% | -0.987% | -1.451% |
| | | (1.075%) | (1.107%) | (1.439%) | (0.818%) | (1.138%) |
| | Trial No No | 0.184% | 0.112% | 0.253% | -7.323% | -0.177% |
| | | (2.492%) | (1.068%) | (1.330%) | (5.327%) | (1.206%) |
| | Trial Trial No | -0.251% | 0.182%* | 0.198% | -3.382% | 0.157% |
| | | (1.949%) | (1.261%) | (1.292%) | (2.006%) | (1.449%) |
| | Trial Trial Trial | -0.043% | 0.610% | 0.172% | 0.683% | -0.052% |
| | | (1.811%) | (1.347%) | (1.234%) | (1.439%) | (1.578%) |
| | Average | 0.864% | 0.579% | 0.384% | -1.046% | -0.235% |
| | | (0.851%) | (0.609%) | (0.664%) | (1.016%) | (0.755%) |
| Membership to Clusters | No Trial No | 2.448%** | 2.041% | 2.463%* | 2.698%† | -0.325% |
| | | (0.915%) | (1.291%) | (1.148%) | (1.549%) | (0.658%) |
| | Trial No No | 1.748% | 0.785% | 2.226%† | 1.503% | 1.505% |
| | | (1.509%) | (1.071%) | (1.234%) | (1.554%) | (1.354%) |
| | Trial Trial No | 1.208% | 0.529% | -0.495% | -1.610% | -1.882% |
| | | (1.125%) | (1.218%) | (1.507%) | (1.549%) | (1.286%) |
| | Trial Trial Trial | -0.041% | 0.782% | 1.023% | 1.343% | 1.701% |
| | | (1.058%) | (1.009%) | (1.084%) | (1.049%) | (1.162%) |
| | Average | 1.313%* | 1.061%† | 1.313%* | 0.845% | 0.625% |
| | | (0.588%) | (0.571%) | (0.618%) | (0.706%) | (0.658%) |

This table reports the difference in the lift in the response rate when using the Timing Model compared to the Base Model (*Increase in Lift*). Both models are used to choose a fixed total of 100 carrier routes to mail. Positive values indicate that the lift in the response rate was higher when using the Timing Model. The lift in the response rate is calculated as the average increase in the response rate in the Treatment condition compared to the Control condition for the carrier routes ranked highest by each model. We omit carrier routes selected by both models. Standard errors are in parentheses.

**Varying the Holdout Sample Size**
**Targeting 100 Carrier Routes: Difference Between Timing and Base Models**

| | | 1 Carrier Route | 5% Holdout | 10% Holdout | 20% Holdout |
|---|---|---|---|---|---|
| Response Before Wave 2 | No Trial No | 2.406% | 0.331% | 2.842%$^{\dagger}$ | 1.806% |
| | | (1.552%) | (1.333%) | (1.674%) | (1.261%) |
| | Trial No No | 2.147%$^{\dagger}$ | 2.573%$^{*}$ | 0.999% | 0.859% |
| | | (1.238%) | (1.244%) | (1.278%) | (1.100%) |
| | Trial Trial No | 0.654% | -0.951% | 0.206% | 0.265% |
| | | (1.506%) | (1.240%) | (1.428%) | (1.062%) |
| | Trial Trial Trial | 1.602% | 1.887% | 0.185% | -0.093% |
| | | (1.174%) | (1.066%) | (0.910%) | (1.060%) |
| | Average | 1.679%$^{**}$ | 1.113%$^{\dagger}$ | 0.818% | 0.633% |
| | | (0.669%) | (0.605%) | (0.626%) | (0.563%) |
| Average Response Date | No Trial No | 1.072% | -0.458% | 0.199% | 2.233% |
| | | (1.439%) | (0.986%) | (1.183%) | (1.510%) |
| | Trial No No | 0.253% | 0.833% | 0.744% | -1.302% |
| | | (1.330%) | (1.339%) | (1.313%) | (1.058%) |
| | Trial Trial No | 0.198% | -0.949% | -1.061% | -0.003% |
| | | (1.292%) | (1.331%) | (1.516%) | (1.240%) |
| | Trial Trial Trial | 0.172% | 1.132% | 0.081% | 0.338% |
| | | (1.234%) | (1.256%) | (1.188%) | (1.220%) |
| | Average | 0.384% | 0.197% | -0.072% | 0.234% |
| | | (0.664%) | (0.645%) | (0.680%) | (0.648%) |
| Membership to Clusters | No Trial No | 2.463%$^{*}$ | 1.496% | 0.739% | 1.467% |
| | | (1.148%) | (1.131%) | (0.980%) | (0.959%) |
| | Trial No No | 2.226%$^{\dagger}$ | 1.696% | 0.970% | 1.473% |
| | | (1.234%) | (1.055%) | (0.972%) | (0.979%) |
| | Trial Trial No | -0.495% | -1.852% | -0.987% | 0.557% |
| | | (1.507%) | (1.481%) | (1.453%) | (1.506%) |
| | Trial Trial Trial | 1.023% | 1.293% | 0.824% | 1.691% |
| | | (1.084%) | (0.991%) | (1.045%) | (1.119%) |
| | Average | 1.313%$^{*}$ | 0.781% | 0.461% | 1.345%$^{*}$ |
| | | (0.618%) | (0.577%) | (0.551%) | (0.566%) |

This table reports the difference in the lift in the response rate when using the Timing Model compared to the Base Model (*Increase in Lift*). Both models are used to choose a fixed total of 100 carrier routes to mail. Positive values indicate that the lift in the response rate was higher when using the Timing Model. The lift in the response rate is calculated as the average increase in the response rate in the Treatment condition compared to the Control condition for the carrier routes ranked highest by each model. We omit carrier routes selected by both models. Standard errors are in parentheses.

**Varying the Minimum Fall Response Threshold: *Response Before Wave 2***

| Minimum Threshold | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 5 or more | No Trial No | 38 | 34 | 0.6374 |
| | Trial No No | 22 | 16 | 0.3304 |
| | Trial Trial No | 20 | 20 | 1.0000 |
| | Trial Trial Trial | 14 | 11 | 0.5485 |
| | Total | 91 | 84 | 0.5967 |
| 8 or more | No Trial No | 30 | 27 | 0.6911 |
| | Trial No No | 39 | 18 | 0.0054** |
| | Trial Trial No | 45 | 39 | 0.5127 |
| | Trial Trial Trial | 45 | 29 | 0.0629† |
| | Total | 159 | 113 | 0.0063** |
| 10 or more | No Trial No | 29 | 36 | 0.3853 |
| | Trial No No | 34 | 19 | 0.0394* |
| | Trial Trial No | 40 | 29 | 0.1854 |
| | Trial Trial Trial | 68 | 35 | 0.0011** |
| | Total | 171 | 119 | 0.0023** |
| 12 or more | No Trial No | 5 | 11 | 0.1336 |
| | Trial No No | 9 | 6 | 0.4386 |
| | Trial Trial No | 25 | 25 | 1.0000 |
| | Trial Trial Trial | 42 | 26 | 0.0523† |
| | Total | 81 | 68 | 0.2869 |
| 15 or more | No Trial No | 12 | 22 | 0.0863† |
| | Trial No No | 11 | 12 | 0.8348 |
| | Trial Trial No | 18 | 16 | 0.7316 |
| | Trial Trial Trial | 38 | 20 | 0.0181* |
| | Total | 79 | 70 | 0.4609 |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Response Before Wave 2* timing variable. In these comparisons we vary the minimum threshold of fall responses used when calculating the *Response Before Wave 2*.

**Varying the Minimum Fall Response Threshold: *Average Response Date***

| Minimum Threshold | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 5 or more | No Trial No | 51 | 52 | 0.9215 |
| | Trial No No | 9 | 7 | 0.6171 |
| | Trial Trial No | 45 | 39 | 0.5127 |
| | Trial Trial Trial | 32 | 30 | 0.7995 |
| | Total | 137 | 128 | 0.5804 |
| 8 or more | No Trial No | 31 | 42 | 0.1979 |
| | Trial No No | 31 | 20 | 0.1235 |
| | Trial Trial No | 65 | 38 | 0.0078** |
| | Trial Trial Trial | 64 | 38 | 0.0100** |
| | Total | 191 | 138 | 0.0041** |
| 10 or more | No Trial No | 16 | 26 | 0.1228 |
| | Trial No No | 19 | 12 | 0.2087 |
| | Trial Trial No | 38 | 33 | 0.9055 |
| | Trial Trial Trial | 63 | 38 | 0.0129* |
| | Total | 136 | 109 | 0.0845† |
| 12 or more | No Trial No | 12 | 12 | 1.0000 |
| | Trial No No | 3 | 5 | 0.4795 |
| | Trial Trial No | 24 | 25 | 0.8864 |
| | Trial Trial Trial | 32 | 24 | 0.2850 |
| | Total | 71 | 66 | 0.6692 |
| 15 or more | No Trial No | 12 | 27 | 0.0163* |
| | Trial No No | 9 | 16 | 0.1615 |
| | Trial Trial No | 14 | 16 | 0.7150 |
| | Trial Trial Trial | 30 | 14 | 0.0159* |
| | Total | 65 | 73 | 0.4959 |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Average Response Date* timing variable. In these comparisons we vary the minimum threshold of fall responses used when calculating the *Average Response Date*.

**Varying the Minimum Fall Response Threshold: *Membership to Cluster***

| Minimum Threshold | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 5 or more | No Trial No | 62 | 56 | 0.5807 |
| | Trial No No | 32 | 24 | 0.2850 |
| | Trial Trial No | 48 | 32 | 0.0736† |
| | Trial Trial Trial | 44 | 35 | 0.3113 |
| | Total | 186 | 147 | 0.0326* |
| 8 or more | No Trial No | 46 | 55 | 0.3705 |
| | Trial No No | 37 | 21 | 0.0356* |
| | Trial Trial No | 40 | 41 | 0.9115 |
| | Trial Trial Trial | 74 | 52 | 0.0500* |
| | Total | 197 | 169 | 0.1433 |
| 10 or more | No Trial No | 42 | 51 | 0.3507 |
| | Trial No No | 21 | 16 | 0.4111 |
| | Trial Trial No | 30 | 28 | 0.7928 |
| | Trial Trial Trial | 82 | 53 | 0.0126* |
| | Total | 175 | 148 | 0.1330 |
| 12 or more | No Trial No | 29 | 37 | 0.3248 |
| | Trial No No | 7 | 7 | 1.0000 |
| | Trial Trial No | 30 | 29 | 0.8964 |
| | Trial Trial Trial | 49 | 38 | 0.2383 |
| | Total | 115 | 111 | 0.7902 |
| 15 or more | No Trial No | 19 | 28 | 0.1893 |
| | Trial No No | 10 | 12 | 0.6698 |
| | Trial Trial No | 20 | 19 | 0.8728 |
| | Trial Trial Trial | 41 | 23 | 0.0244* |
| | Total | 90 | 82 | 0.5419 |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Membership to Clusters* timing variables. In these comparisons we vary the minimum threshold of fall responses used when estimating the *Membership to Cluster* variables.

**Varying the Holdout Sample Size: *Response Before Wave 2***

| | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 1 Carrier Route Holdout | No Trial No | 29 | 36 | 0.3853 |
| | Trial No No | 34 | 19 | 0.0394* |
| | Trial Trial No | 40 | 29 | 0.1854 |
| | Trial Trial Trial | 68 | 35 | 0.0011** |
| | Total | 171 | 119 | 0.0023** |
| 5% Holdout | No Trial No | 35 | 39 | 0.6419 |
| | Trial No No | 35 | 21 | 0.0614† |
| | Trial Trial No | 42 | 31 | 0.1979 |
| | Trial Trial Trial | 69 | 37 | 0.0019** |
| | Total | 181 | 128 | 0.0026** |
| 10% Holdout | No Trial No | 36 | 41 | 0.5688 |
| | Trial No No | 26 | 22 | 0.5637 |
| | Trial Trial No | 48 | 29 | 0.0304* |
| | Trial Trial Trial | 66 | 33 | 0.0009** |
| | Total | 176 | 125 | 0.0033** |
| 20% Holdout | No Trial No | 31 | 41 | 0.2386 |
| | Trial No No | 35 | 18 | 0.0195* |
| | Trial Trial No | 45 | 32 | 0.1385 |
| | Trial Trial Trial | 64 | 31 | 0.0007** |
| | Total | 175 | 122 | 0.0021** |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Response Before Wave 2* timing variable. In these comparisons we cross validate by holding out different proportions of the 796 carrier routes in each cross validation.

**Varying the Holdout Sample Size: *Average Response Date***

| | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 1 Carrier Route Holdout | No Trial No | 16 | 26 | 0.1228 |
| | Trial No No | 19 | 12 | 0.2087 |
| | Trial Trial No | 38 | 33 | 0.9055 |
| | Trial Trial Trial | 63 | 38 | 0.0129* |
| | Total | 136 | 109 | 0.0845† |
| 5% Holdout | No Trial No | 25 | 29 | 0.5862 |
| | Trial No No | 21 | 15 | 0.3173 |
| | Trial Trial No | 42 | 34 | 0.3588 |
| | Trial Trial Trial | 66 | 39 | 0.0084** |
| | Total | 154 | 117 | 0.0246* |
| 10% Holdout | No Trial No | 23 | 34 | 0.1451 |
| | Trial No No | 17 | 12 | 0.3532 |
| | Trial Trial No | 41 | 30 | 0.1917 |
| | Trial Trial Trial | 60 | 34 | 0.0073** |
| | Total | 141 | 110 | 0.0504† |
| 20% Holdout | No Trial No | 25 | 37 | 0.1275 |
| | Trial No No | 28 | 18 | 0.1404 |
| | Trial Trial No | 42 | 40 | 0.8252 |
| | Trial Trial Trial | 63 | 29 | 0.0004** |
| | Total | 158 | 124 | 0.0429* |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Average Response Date* timing variable. In these comparisons we cross validate by holding out different proportions of the 796 carrier routes in each cross validation.

**Varying the Holdout Sample Size: *Membership to Cluster***

| | Treatment | Timing Model | Base Model | P-value |
|---|---|---|---|---|
| 1 Carrier Route Holdout | No Trial No | 42 | 51 | 0.3507 |
| | Trial No No | 21 | 16 | 0.4111 |
| | Trial Trial No | 30 | 28 | 0.7928 |
| | Trial Trial Trial | 82 | 53 | 0.0126* |
| | Total | 175 | 148 | 0.1330 |
| 5% Holdout | No Trial No | 43 | 50 | 0.4679 |
| | Trial No No | 23 | 11 | 0.0396* |
| | Trial Trial No | 30 | 35 | 0. 5351 |
| | Trial Trial Trial | 78 | 53 | 0.0289** |
| | Total | 179 | 149 | 0.1642 |
| 10% Holdout | No Trial No | 45 | 44 | 0.9156 |
| | Trial No No | 27 | 16 | 0.5327 |
| | Trial Trial No | 36 | 28 | 0.3173 |
| | Trial Trial Trial | 80 | 56 | 0.0396* |
| | Total | 188 | 144 | 0.0157* |
| 20% Holdout | No Trial No | 44 | 53 | 0.3608 |
| | Trial No No | 34 | 11 | 0.0006** |
| | Trial Trial No | 39 | 33 | 0.4795 |
| | Trial Trial Trial | 73 | 49 | 0.0298* |
| | Total | 190 | 146 | 0.0164* |

The table reports the number of carrier routes in which the Timing Model and Base Model make the correct targeting recommendation. The Timing Model uses the *Membership to Clusters* timing variables. In these comparisons we cross validate by holding out different proportions of the 796 carrier routes in each cross validation.