# Using the Timing of Past Responses in Targeting Models to Address Dilution

April 2020

Theodoros Evgeniou
theodoros.evgeniou@insead.edu

Duncan Simester
simester@mit.edu

Artem Timoshenko
artem.timoshenko@kellogg.northwestern.edu

Spyros I. Zoumpoulis
spyros.zoumpoulis@insead.edu

Dilution of a market can hinder the performance of models designed to target different customers with different marketing actions. We propose using the timing of past responses as a measure of dilution. Using data from two field experiments, we show that including information about the timing of past responses can significantly improve the performance of targeting policies. The timing of past response is particularly valuable for targeting future firm actions when there is more variation in the effectiveness of those actions.

## 1. Introduction and Related Work

As marketing decisions are increasingly based on data, a key challenge that managers face is to adjust for any changes in the business environment that may make past data misleading. This might include changes in customer tastes or interests, introduction of new products, or shifts in macro conditions. Data-driven decision making requires that firms consider potential changes over time, or risk that their models recommend suboptimal decisions. This risk is amplified when there is a material delay between the collection of data and the implementation of policies based upon it.

One important source of change is dilution of a market. Dilution occurs in contexts in which customers can only respond to marketing actions once, so those who responded in the past cannot respond again in the future. For example, when prospecting for new customers, those who have adopted in the past are no longer considered prospects, and so are typically removed from future prospecting pools. Similarly, when reactivating lapsed customers, those that are reactivated are no longer lapsed. Dilution may also arise when cross-selling to existing customers. For example, once life insurance customers have been cross-sold car insurance, the opportunities to sell additional car insurance to the same customer pool becomes diluted.

We focus on the information that past responses provide about future responses, and how this information is affected by dilution. Consider a prospecting example, in which a firm is choosing which zip codes to mail a promotion. Is a zip code that has yielded many new customers in prior periods also likely to yield many customers in the future? If the high past response indicates that the preferences of these customers are well suited to the promotion, this argues for a higher future response. However, if the high past response leads to fewer remaining good prospects, this argues for a lower future response. We label this second effect “dilution”, and study how dilution in a market affects firm decisions.

We motivate our empirical work using an illustrative model. The model explains why the timing of past responses can provide a measure of how diluted a market has become. The model also illustrates why we expect larger future responses from markets that are less diluted, and why dilution plays a more important role when the firm’s marketing actions are more effective. This last observation is a key feature of our empirical work.

The empirical work uses (secondary) data from two field experiments implemented by a large chain of warehouse clubs. It demonstrates that timing information is useful not just when predicting future responses, but also when measuring how future responses vary with changes in marketing actions. We show that in zip codes with little decline in past responses over time, a large number of past responses is a strong signal that the future responses will also be high. In contrast, in zip codes that have had a sharp decline in responses over time, a large number of past responses is a weaker (less positive) signal of future responses. We also confirm that the role of the timing information depends upon the effectiveness of the firm’s future marketing actions.

We illustrate how to incorporate the timing of past responses when training a targeting policy. Using data from field experiments, we demonstrate that a targeting policy trained using the timing of past responses can significantly outperform a policy trained without using timing. The size of the performance improvement depends upon the firm’s action space. The timing of past response is especially useful for targeting future actions when there is more variation in the effectiveness of those actions. This is not simply because it is easier to make targeting decisions when there is more variation in the actions' effectiveness. We compare the performance of a model that incorporates timing information, with a benchmark model that does not, and so any features that make a problem "easier" should benefit both models. Instead, the performance improvement is magnified because the effectiveness of a firm's marketing actions determines the value of timing information.

### 1.1 Related Work

We study dilution as a source of non-stationarity in targeting models. Marketing research has documented the importance of non-stationarity and proposed methods to address it in various contexts (Fader and Lattin, 1993). For example, Naik, Mantrala, and Sawyer (1998) study wearin and wearout effects and estimate a dynamic model of the effect of promotion repetition on customer awareness. Netzer, Lattin, and Srinivasan (2008) estimate a hidden Markov model to infer customer transitions between latent relationship states and their effects on purchasing behavior. Chae, Bruno, and Feinberg (2019) develop a model that captures varying degrees of wearout to improve ad deployment. Zhang (2020) provides a recent review of the literature on customer dynamics and the empirical approaches to understanding these dynamic. Our scope is different from this strand of marketing research because we focus on dilution. Dilution describes changes in the composition of the customer pool over time, rather than changes in the individual response behavior. We show that these changes can be anticipated and measured.

Dilution is relevant when individual customers purchase a single time. Past adoptions result in dilution of the pool, unless the target customer pool refreshes quickly. Practical examples include prospecting (Schwartz, Bradlow, and Fader, 2017; Simester, Timoshenko, and Zoumpoulis, 2019a), reactivating lapsed customers (Ascarza, 2017), and cross-selling of individual products (Gurvich, Armony, and Maglaras, 2009). We focus on customer acquisition (prospecting). Firms generally do not have any prior purchase history for prospective customers and so they are forced to rely upon demographic or geographic data (Simester et al., 2019a). 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. Although our data describes a firm's actions when prospecting for new customers, we expect that our results will generalize beyond prospecting to include applications in both cross-selling and reactivating lapsed customers.

We can also compare our findings with the extensive literature on customer lifetime value (CLV).[1] The CLV model is a model of individual customer behavior. This contrasts with dilution, which describes aggregate behavior in a market. However, all of these theories share a common feature: they anticipate changes in customer responses over time. A key parameter in the CLV model is the churn rate, which recognizes that some customers will end their relationship with the company each period. Future profits are only earned from customers who did not churn, and so in some sense, the pool of customers becomes diluted over time. An obvious difference between dilution in a CLV model and dilution in this

[1] See for example: Kumar (2008) and Fader, Hardie and Lee (2005).

paper is that CLV captures dilution in existing customer relationships, while we measure dilution in the customer acquisition process. However, even though it focuses on dynamics in existing customer relationships, CLV is an important concept when prospecting for new customers. Because the CLV model can be used to calculate the discounted future net profit earned from a customer, it provides an upper bound on how much a firm should spend to acquire that customer (Farris et al. 2010).

Beyond dilution, other research has leveraged timing information to improve the performance of targeting models. For example, Rafieian and Yoganarasimhan (2020) create targeting variables by summarizing contextual, behavioral and ad-related information over long-term, short-term and session-level histories. Negoescu et al. (2018) show that the timing of infrequent health events can help to allocate treatments to patients more optimally. Because neither of these studies consider the role of dilution, they do not consider how the information that we learn from the timing of past responses depends upon the effectiveness of the targeted actions.

By demonstrating the value of timing information for targeting, our paper relates to research that has sought to improve targeting models by introducing additional information and features. 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) highlight 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.

### 1.2 Structure of the Paper

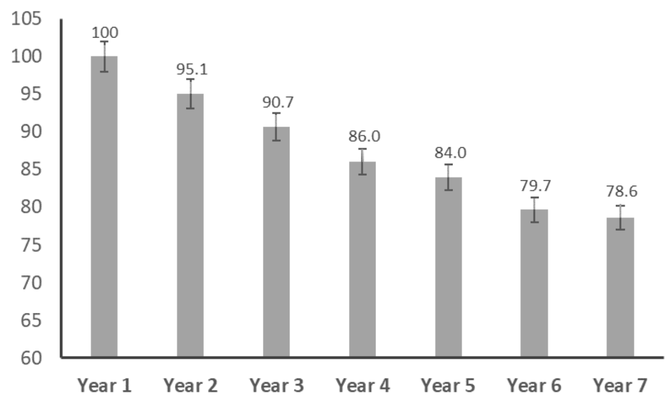
Section 2 highlights the importance of dilution, and Section 3 describes an illustrative model. These sections are designed to motivate and help interpret the analysis that follows. In Section 4, we use data from two field experiments to investigate the relationship between past and future responses. We show that the relationship depends upon the timing of past responses, and also investigate how the effectiveness of the firm's marketing actions affects the value of the timing information. Section 5 extends these results to targeting models, and demonstrates how to incorporate timing information when training a targeting model. The paper concludes in Section 6.

## 2. The Importance of Dilution

In Figure 1 we report the aggregate number of new customers acquired each year over a seven-year period by a large chain of warehouse clubs. We focus on customers acquired from a fixed sample of zip codes, by restricting attention to zip codes that contributed at least one new customer in each of the seven years. For confidentiality reasons the numbers are indexed to 100 in year 1.[2]

[2] We also do not report the exact dates of the period. Customers have to sign up for memberships in order to shop in the stores, and for each customer the acquisition date reflects the date that the customer first signed up for a membership.

**Figure 1. Total New Customer Adoptions by Year**




The figure reports a count of the number of new customers acquired by a large warehouse retailer over a seven-year period. The counts are indexed to 100 in year 1. Error bars indicate 95% confidence intervals.

The average number of new customers acquired within each zip code falls monotonically across the seven years. In the seventh year, the firm acquires 21.4% fewer customers than it did in the first year. Customers represented in this figure comprise the retailer's pipeline of new customers. Figure 1 reveals that acquiring new customers is becoming substantially more difficult over time. This pattern has dramatic implications for the retailer, both economically, as well as for decision-making going forward.

The trend in Figure 1 is consistent with dilution of the pool of prospective customers. In later years, customers with the highest probability of adopting have already chosen to do so earlier, and so the remaining customers that have not yet joined are an increasingly pure pool of households that have a low probability of adopting. This is the phenomenon that we study.

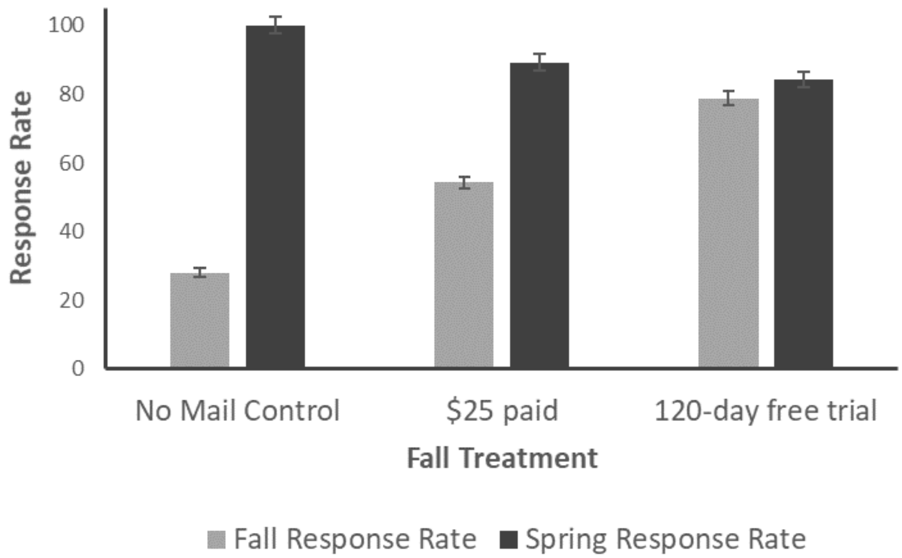
However, other factors may also have contributed to the pattern, such as a decline in the firm's marketing budget over time. For this reason, we illustrate the importance of dilution in a more controlled example using two field experiments conducted by the same wholesale club. The experiments sent promotions to prospective customers and were intended to test the effectiveness of different offers designed to attract new members.

**Example: Results of Two Field Experiments**

The two experiments were conducted approximately six months apart, in Fall 2015 and Spring 2016. In the Fall 2015 experiment, households were randomly assigned to one of three experimental conditions: a $25 discounted membership, a 120-day free trial, and a no mail control. In the Spring 2016 experiment, the retailer repeated the same three experimental conditions using the same households. However, it re-randomized, so that households in Spring 2016 were equally likely to receive any of the promotions. Notably, households that responded in Fall 2015 were removed from the Spring 2016 experiment.

In Figure 2 we report the average response rate for each type of promotion, when grouping the households according to the promotion they received in the Fall.

**Figure 2. Comparison of Fall and Spring Response Rates**




The figure reports the average Fall and Spring response rates, grouped according to the treatment received in Fall. The response rates are indexed to 100 in the Spring No Mail condition. Error bars indicate 95% confidence intervals. We provide additional details about the two experiments in the Appendix.

Households that were in the Fall treatment with the highest Fall response rate (the 120-day free trial Fall condition) had the lowest average response rate in Spring. In contrast, households in the Fall *No Mail Control* had the lowest response rate in Fall, but the highest response rate in Spring. We know that the Fall treatments caused the variation in the Spring response rates.[3] We can conclude that the underlying response rate changed systematically from Fall to Spring. Randomization ensured that the three groups of households were equivalent at the start of Fall, but they were no longer equivalent at the start of Spring. This is an example of how a market can change over time.

The shift in response rates between Fall and Spring is consistent with dilution. The households that responded in Fall were removed from the Spring prospect pool. As a result, the households that received the 120-day free trial in Fall had fewer remaining potential responders in Spring. In the presence of dilution, a large past response (in Fall) may signal a *low* future response (in Spring). It is this type of systematic changes in response rates over time that we seek to identify and incorporate in a targeting model.[4]

[3] Prior to the Fall treatments these groups of households were equivalent (the Fall treatments were randomly assigned). Because the treatments were randomized again in Spring, each of the three groups received the same distribution of treatments in Spring.

[4] It is important to recognize that past responses can also signal information about underlying customer preferences. When this occurs, the relationship between past and future responses may be reversed; a high past

It is again possible that other factors may have contributed to the findings in Figure 2. In Sections 4 and 5, we use the data from two experiments to further investigate the information that past responses provide about future responses. We show how this relationship depends upon the timing of the past responses, and demonstrate that the findings vary with the effectiveness of the firm's marketing actions. These findings illustrate relatively sophisticated interactions, and so are much less susceptible to alternative explanations.

## 3. An Illustrative Model of Dilution for Targeting

We present a simple illustrative model to help understand how dilution and the effectiveness of the firm's marketing actions interact to shape future responses. The model is designed to merely illustrate and clarify, rather than to offer a theoretical contribution.

We consider a firm engaged in customer acquisition promotions conducted in three stages. The firm needs to decide which markets to send a promotion at Stage 3 after having gathered data from Stages 1 and 2.

We model each market (e.g., a zip code) as having two types of customers: Type R customers can potentially respond to a promotion, and Type G customers never respond. There are $r$ Type R and $g$ Type G customers before Stage 1, so the total number of customers initially in the market is $n = r + g$. In each of Stages 1 and 2, exactly $q^{past}$ proportion of the remaining Type R customers respond. We interpret $q^{past}$ as a parameter that controls dilution: a high $q^{past}$ means many of the Type R customers respond in Stages 1 and 2, leaving only a few possible responders in Stage 3.

The firm's marketing actions in Stage 3 can change the proportion of remaining Type R customers that respond. We use $q^{action}$ to describe the proportion of remaining Type R customers that respond in Stage 3, conditional on the firm implementing the treatment denoted by $action$.

In Stage 1, $s_1$ Type R customers respond and are removed from the market. At the start of Stage 2 there are $r - s_1$ Type R customers and $g$ Type G customers, for a total of $n - s_1$ customers remaining. In Stage 2, an additional $s_2$ Type R customers respond and are also removed from the market. As a result, at the start of Stage 3 there are $r - s_1 - s_2$ Type R and $g$ Type G customers left. We assume that the firm observes the values of $n$, $s_1$, and $s_2$ for each market. Parameters $r$ and $q^{past}$ are unknown *a priori* for any market, and parameters $n, r, g, q^{past}$, and $q^{action}$ can differ across markets.[5]

Given these definitions, we can describe the number of Stage 1 and Stage 2 responses, as well as how the number of Stage 3 responses will change with the firm's marketing actions: [6]

---

response may signal a higher future response. In Figure 2 this preference effect is removed because there is no systematic difference in customer preferences across the three groups. The households were randomly assigned to the Fall treatments, and so there is no variation in preferences to identify.

[5] The number of Type G customers $g$ is also unknown, but if we learn $r$ we also learn $g$.

[6] We could also allow $s_1$ and $s_2$ to vary with the firm's past marketing actions. However, in the empirical analysis that follows we do not observe the firm's historical marketing actions, and so we use a generic $q^{past}$ for these earlier stages of the model.

$$s_1 = r\, q^{past}$$

$$s_2 = (r - s_1)q^{past} = (r - rq^{past})q^{past} = rq^{past}(1 - q^{past})$$

The goal is to learn (predict) the number of customers that will respond in Stage 3 ($s_3$) so that we can then select which markets to target:

$$s_3 = (r - s_1 - s_2)q^{action} = [r - rq^{past} - rq^{past}(1 - q^{past})]q^{action} = rq^{action}(1 - q^{past})^2$$

**Observation**: The response in Stage 3, $s_3$, is monotonically increasing in $r$ and $q^{action}$, and monotonically decreasing in $q^{past}$.

We consider two cases. One case is that we only observe (or consider) the *Total Past Response* ($s_1 + s_2$) but we do not observe $s_1$ and $s_2$ separately. In this case we do not know the timing of past responses (Stage 1 versus Stage 2). The second case is that we observe and use $s_1$ and $s_2$ individually.

In the first case, we cannot identify $r$ and $q^{past}$, as we have only one piece of information and two unknowns. On the other hand, if we know both $s_1$ and $s_2$, we can identify market parameters $r$ and $q^{past}$ using equations for $s_1$ and $s_2$:

$$q^{past} = \frac{s_1 - s_2}{s_1} \text{ (or } 1 - q^{past} = \frac{s_2}{s_1}\text{)}$$

$$r = \frac{s_1^2}{s_1 - s_2}$$

We can then write $s_3 = rq^{action}\left(\frac{s_2}{s_1}\right)^2$. The response in Stage 3 is larger when:

- The market begins before Stage 1 with more potential responders (Type R customers): $r$ is large.
- The future marketing action is more effective: $q^{action}$ is large.
- Past responses have fallen more slowly over time, indicating less dilution: $q^{past}$ is small, or equivalently, $s_2/s_1$ is large.

This expression highlights the role of the timing of past responses. A larger share $s_2/s_1$ signals that the market is less diluted at the start of Stage 3. When responses have decreased more slowly over time, there are more good prospects remaining in the market at the end of Stage 2. We also see that the effectiveness of a firm's future marketing actions regulates this relationship. When future marketing actions are more effective ($q^{action}$ is larger), the timing of past responses plays a more important role. This is an important observation, which will motivate key aspects of our analysis in the next two sections.

We finish this section with two comments. Our focus is on the change in the composition of Type R and Type G customers in a market over time, which is how we interpret the role of dilution. We can contrast this with changes in the response probabilities for an individual customer over time. For example, other researchers have focused on the role of wearin or wearout of a promotion over time (Pechmann and Stewart 1988; Chae, Bruno, and Feinberg, 2019). While dilution is reflected in the ratio of $s_1$ and $s_2$, wearin and wearout would be reflected in differences between $q^{past}$ and $q^{action}$.

Our second comment concerns the distinction between responses and response rates. In many applications, firms focus on response rates to explicitly control for the number of households in different markets (as we will do when training a targeting model in Section 5). We can alternatively present the outcome in Stage 3 as a response rate (assuming $g > 0$):

$$\frac{s_3}{n - s_1 - s_2} = \frac{rq^{action}(1 - q^{past})^2}{g + r - r\,q^{past} - rq^{past}(1 - q^{past})} = \frac{rq^{action}(1 - q^{past})^2}{g + r(1 - q^{past})^2}$$

Although this expression is more complicated, the main insight of the model for the outcome in Stage 3 remains the same. Like the response, the response rate is increasing in $r$, $q^{action}$ and decreasing in $q^{past}$. Moreover, the role of dilution ($q^{past}$) is again moderated by the effectiveness of the firm's future marketing actions ($q^{action}$). In the next section we will use a sample of customer acquisition data to empirically investigate these relationships.

**Summary**

We have presented a model to illustrate the intuition behind the findings that we will present in the next two sections. The model yields three important observations:

1. The timing of past response provides a measure of how diluted a market has become. This information is not available from just observing the aggregate past response.
2. The less diluted a market, the larger the response in the next stage.
3. Dilution plays a more important role when the firm's future marketing actions are more effective.

## 4. The Timing of Past Responses and the Response to Future Marketing Actions

We use data from two field experiments to study the relationship between past and future responses, the timing of past responses, and the effectiveness of a firm's marketing actions. The two experiments were both conducted by the same retailer that provided the data used in Section 2. They were conducted in Spring 2015 and Spring 2016, approximately 12 months apart (we provide details about the experiments in the Appendix).[7]

In both experiments, the retailer randomly varied the same three experimental treatments that we described in Section 3: a $25 discounted membership, a 120-day free trial, and a No Mail control. The randomization of the three experimental conditions was conducted at the household level. In each experiment, every zip code contained an equivalent subsample of households that received the 120-day treatment, a subsample that received the $25 treatment, and a subsample that did not receive either

[7] The Spring 2016 experiment was the same experiment that we described in Figure 2. We do not use the Fall 2015 experiment that was used in Figure 2, because in that experiment the variation was at the carrier route level, not the household level. As a result, we do not have equivalent groups of customers in each zip code that received each experimental treatment (there are an average of only 13 carrier routes in each zip code). Both experiments also included other experimental conditions (see the Appendix).

treatment (the No Mail control). The Spring 2015 experiment was conducted in two regions, while the Spring 2016 was conducted across a much larger set of regions. After omitting zip codes for which we have no demographic data, the Spring 2015 experiment yields data for 412 zip codes, while the Spring 2016 experiment includes 2,558 zip codes.

We first analyze the results when aggregating across all of the households in a zip code, irrespective of which of the three treatments they received.[8] We then repeat the analysis separately for each experimental treatment.

**Aggregate Analysis (Aggregating Across Treatments)**

Recall our discussion in the Introduction. We argued that the more sharply the rate of response to promotions declines over time, the more diluted a market has become. For a market with little decline in the response rate over time, a large number of past responses is a strong signal that the future responses will also be high. In contrast, if a market has had a sharp decline in the response rate over time, a large number of past responses is a weaker (less positive) signal of future responses. Our aggregate analysis focuses on investigating this relationship. In particular, we estimate the following OLS models:

**Responses**

$$Response_z = \alpha + \beta_1\, Past\ Response_z * Timing_z + \beta_2\, Past\ Response_z + \beta_3\, Timing_z + \gamma\, Demographics_z + \varepsilon_z \qquad (1)$$

**Response Rates**

$$Response\ Rate_z = \alpha + \beta_1\, Past\ Response\ Rate_z * Timing_z + \beta_2\, Past\ Response\ Rate_z + \beta_3\, Timing_z + \gamma\, Demographics_z + \varepsilon_z \qquad (2)$$

The dependent variable in Equation 1 is a count of the responses received in zip code *z* (in that experiment). In Equation 2, a response rate is calculated by dividing this count by the total number of households in zip code *z* that participated in the experiment. The model estimates how the interaction between *Past Response* (or *Past Response Rate*) and *Timing* contribute to each of these outcomes.

We measure the *Past Response* in a zip code using the number of members acquired from that zip code in the prior calendar year (January – December). For the Spring 2015 experiment, this is the number of responses in 2014. We do not observe the historical marketing actions in each zip code, and so this measure of past response does not consider the firm's historical marketing actions (see also Footnote 6). We calculate the *Past Response Rate* by dividing the *Past Response* measure by a count of the number of households in the zip code (this count was provided by the retailer).

The timing of past responses (*Timing*) measures the proportion of *Past Responses* in the prior calendar year that occurred in the second half of that prior year (July – December). Intuitively, this measures the extent to which the response across the prior year was increasing or decreasing. We expect $Timing_z$ to measure dilution (together with other forms of non-stationarity); the faster a zip code becomes diluted,

[8] In this initial analysis we exploit the randomization of the treatments to the extent that it allows us to claim that the distribution of the treatments was not systematically different across the zip codes.

the smaller this ratio will be. We recognize that there are other ways to capture the timing of past responses. As a robustness check, we repeated the analysis using different periods to measure *Past Response* and *Timing*. This yielded a similar pattern of findings.[9]

*Demographics* refers to a set of nine zip-code level demographic variables that the firm provided for the 412 zip codes in the Spring 2015 experiment and the 2,558 zip codes in the Spring 2016 experiment. Definitions and summary statistics for these demographic variables are provided in the Appendix.

The coefficient of interest in both models is $\beta_1$, which measures how the interaction between *Past Response* (or *Past Response Rate*) and *Timing* contributes to future responses. If the *Timing* ratio is large, indicating less dilution, we expect that past responses will be a stronger signal of future response. This interpretation argues that this interaction coefficient should be positive. We estimate the two models separately for each experiment (Spring 2015 or Spring 2016). The sample sizes are 412 (Spring 2015) and 2,558 (Spring 2016). The observations in the response rate models are weighted by the number of households in the zip code that participated in the experiment (the weighting is implicit in the response models). The coefficients of interest are reported in Table 1.

**Table 1. Timing and the Relationship Between Past and Future Responses**

| | **Responses** | | **Response Rates** | |
|---|---|---|---|---|
| **Experimental Condition** | **Spring 2015** | **Spring 2016** | **Spring 2015** | **Spring 2016** |
| *Past Response (Rate) * Timing* | 0.1549** <br> (0.0275) | 0.0029** <br> (0.0006) | 0.3034** <br> (0.0713) | 0.0586† <br> (0.0350) |
| $R^2$ | 0.9213 | 0.2511 | 0.8363 | 0.3773 |

The table reports the *Past Response (Rate) * Timing* coefficients ($\beta_1$) from estimating Equations 1 and 2. The unit of analysis is an experiment (Spring 2015 or Spring 2016) x zip code. The sample sizes are 412 (Spring 2015) and 2,558 (Spring 2016). Standard errors are in parentheses. The observations in the response rate models are weighted by the number of households in the zip code (that participated in the experiment). **Indicates significantly different from zero ($p < 0.01$). *Indicates significantly different from zero ($p < 0.05$). †Indicates significantly different from zero ($p < 0.10$).

Differences in the implementation of the experiments mean that there are differences in the magnitudes of the Spring 2015 and Spring 2016 coefficients (particularly for the *Response* models). However, this just reflects differences in the implementation of the two experiments (see the Appendix). Instead, in this analysis we are focused on the sign of the interaction coefficients. As expected, in all four models the interaction coefficient is positive, which is consistent with our

[9] Although the model in Section 3 was intended for illustration and clarification, and we did not intend to directly estimate it, the interactions in Equations 1 and 2 are consistent with that model. If we define $s_1$ as the start of the prior year, and $s_2$ as the end of the prior year, then we can write *Past Responses* = $s_1 + s_2$ and *Timing* = $\frac{s_2}{s_1+s_2}$. Substituting and re-arranging yields: $s_3 = q^{action} \cdot Past\ Response \cdot \frac{Timing^2}{1-2\ Timing}$. We see that the interaction between *Past Responses* and *Timing* contributes to future responses ($s_3$).

interpretation that past responses are a stronger indicator of future responses when there is less evidence of dilution. In contrast, when there is a lot of dilution, indicated by a larger decrease in past responses over time, then past responses become a weaker signal of future responses.

We can illustrate this relationship more clearly by grouping the observations using a median split of the *Timing* measure and estimating the following equations separately within each subsample:

$$Response_z = \alpha + \beta_1\ Past\ Response_z + \gamma\ Demographics_z + \varepsilon_z \tag{3}$$

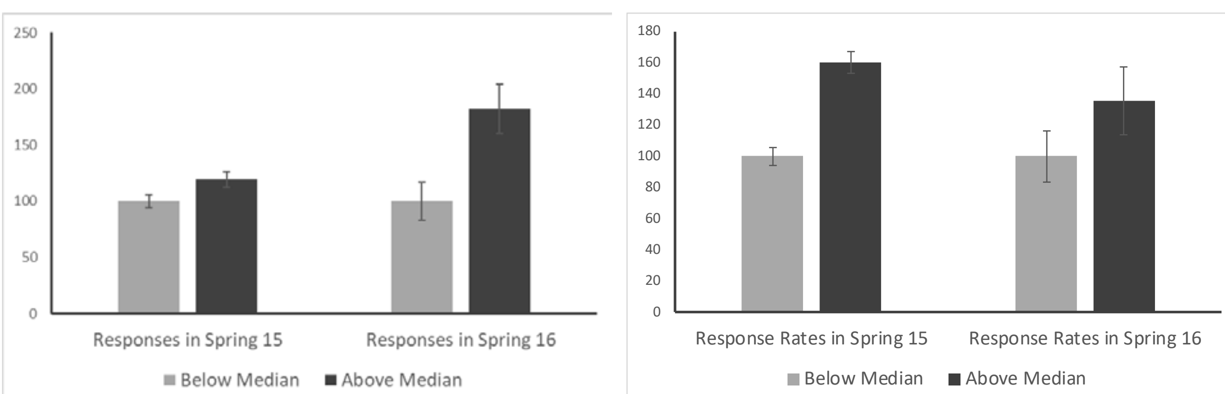
$$Response\ Rate_z = \alpha + \beta_1\ Past\ Response\ Rate_z + \gamma\ Demographics_z + \varepsilon_z \tag{4}$$

In Figure 3 we summarize the *Past Response* or *Past Response Rate* coefficients ($\beta_1$) when re-estimating the models separately using the subsamples created by each median split. To aid interpretation, we index the coefficients estimated in the below-median subsample to 100. We see that when the *Timing* measure is smaller (indicating greater dilution), the *Past Response* coefficient is significantly smaller.

Recall again our discussion in the Introduction. We considered a zip code with a high past response and asked how this would affect a ranking of which zip codes are expected to have the highest future response. We recognized that if the high past response indicates that customer preferences are well-suited to the promotion, this argues for a higher future response. In contrast, if the high past response leads to fewer potential future responders, this argues for a lower future response. Figure 3 illustrates that the weight we should give to past response varies with the rate of dilution. Zip codes that have a higher rate of dilution (lower *Timing* values) should give lower weight to past responses than zip codes that have a lower rate of dilution.

**Figure 3. Timing and the Relationship Between Past and Future Responses**

**Responses** **Response Rates**




The figure reports the coefficients ($\beta_1$) from estimating Equations 3 and 4 separately using median splits of the *Timing* measure. Each coefficient is obtained from a separate model. Error bars indicate 95% confidence intervals. Additional details (including sample sizes) are reported in the Appendix.

As a robustness check, we repeated this analysis using the customer acquisition data provided by the retailer. We used this file in Figure 1 to show that the response rate monotonically decreased over a seven-year period. In the Appendix we use each zip code's 2014 and 2015 responses to predict that zip code's 2016 responses. This robustness check again confirms that the relationship between past and future responses depends upon the timing of the past responses. What makes this particularly reassuring is that this data represents the retailer's entire population of new customers. The more slowly that responses have decreased over time, the more positive the relationship between past and future responses. In our next analysis, we investigate how this relationship is affected by the effectiveness of the firm's marketing actions.

**The Effectiveness of the Firm's Marketing Actions**

We have shown that the relationship between (future) *Response* and *Past Response* depends upon the *Timing* of the past responses. However, in our illustrative model, we also demonstrated that the importance of *Timing* varies with the effectiveness of the firm's future marketing actions. When the future actions are more effective, *Timing* (and thus dilution) plays a more important role. This is a three-way interaction, which can be challenging to interpret. Intuitively, if the future marketing action is ineffective so that few customers respond, then the number of good prospects remaining in the zip code is relatively less important. However, if the marketing action is very effective, so that a high proportion of good prospects are likely to respond ($q^{action}$ is large), then the extent to which the market is diluted becomes very important.

The retailer's Spring 2015 and Spring 2016 experiments provide an opportunity to investigate this issue. In each experiment we can separate households in each zip code into equivalent sub-groups according to the experimental treatment that the household was randomly assigned to receive. We then re-estimate Equations 1 and 2 separately in each sub-group. The *Past Response (Rate) * Timing* coefficients are reported in Table 2 ($R^2$ values are reported in the Appendix).

We again stress that the comparison between the 2015 and 2016 coefficients is not the important comparison (this just captures differences in the implementation of the two experiments). Instead the key comparison is how the coefficients vary across the three experimental treatments. We expect that the interaction between *Timing* and *Past Response* will be more positive when the marketing action is more important. To evaluate this prediction, we also need a measure of the effectiveness of the three experimental treatments. Fortunately, the experiments provide a relevant measure; we measure the effectiveness of the treatments by the average response rate in each experimental condition.[10]

In Figure 4 we report the average response rate under each marketing action and compare this with the size of the associated interaction coefficient ($\beta_1$). The light bars report the average response rate under

[10] According to our model, if we aggregate response rates across zip codes (within a treatment) we obtain: $\sum_z \frac{s_{3,z}^{action}}{n_z - s_{1,z} - s_{2,z}} = q^{action} \sum_z \frac{r_z(1-q^{past})^2}{g_z + r_z(1-q^{past})^2}$. Notice that $\sum_z \frac{r_z(1-q^{past})^2}{g_z + r_z(1-q^{past})^2}$ does not vary across the three marketing actions, and so we can use the empirical average response rate in each treatment as a relative measure of the effectiveness of each marketing action ($q^{action}$).

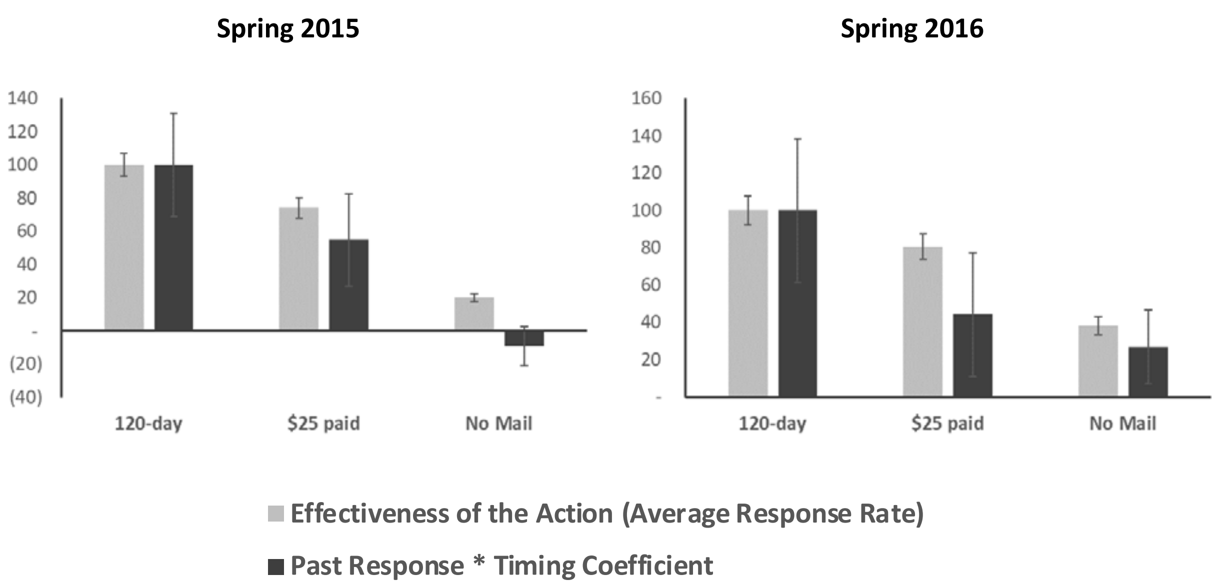
each treatment. The dark bars depict the *Past Response* * *Timing* ($\beta_1$) coefficients reported in Table 2. For ease of comparison, both measures are indexed to 100 in the 120-day condition.

**Table 2. Results by Experimental Condition**

| | Responses | | Response Rates | |
|---|---|---|---|---|
| **Experimental Condition** | **Spring 2015** | **Spring 2016** | **Spring 2015** | **Spring 2016** |
| 120-day Trial | 0.3193** | 0.0052** | 0.6617** | 0.2273** |
| | (0.0508) | (0.0010) | (0.1356) | (0.0719) |
| $25 Paid | 0.1756** | 0.0023* | 0.2306† | -0.0300 |
| | (0.0455) | (0.0009) | (0.1306) | (0.0599) |
| No Mail | -0.0294 | 0.0014** | 0.0225 | 0.0028 |
| | (0.0192) | (0.0005) | (0.0577) | (0.0395) |

The table reports the *Past Response (Rate)* * *Timing* coefficients ($\beta_1$) from estimating Equations 1 and 2. Each coefficient is estimated using a separate model. The unit of analysis is an experiment (Spring 2015 or Spring 2016) x zip code x marketing action (120-day, $25, or No Mail). The sample sizes are 412 (Spring 2015) and 2,558 (Spring 2016). Standard errors are in parentheses. The observations in the response rate models are weighted by the number of households in the zip code (that participated in the experiment). **Indicates significantly different from zero (p < 0.01). *Indicates significantly different from zero (p < 0.05). †Indicates significantly different from zero (p < 0.10).

**Figure 4. The Importance of Timing Depends Upon the Effectiveness of the Marketing Actions**




The light bars report the average response rate across zip codes in each experimental condition. The dark bars report the *Past Response* * *Timing* coefficient ($\beta_1$) from estimating Equation 1 separately for each experimental condition. Both measures are indexed to 100 in the 120-day condition.

The light bars indicate that the 120-day trial was the most effective marketing treatment, followed by the $25 paid promotion, and then the No Mail control condition. The relative effectiveness of the three treatments, measured by the (indexed) average response rate, is notably consistent across the two experiments.

Given these results, our illustrative model predicts that dilution will play a more important role in the 120-day trial condition. The findings of both experiments are consistent with this. In both experiments, the ordering of the *Past Response* * *Timing* coefficient across the three marketing actions (the dark bars) is consistent with the ordering of the average effectiveness of each marketing action (the light bars). The 120-day trial was the most effective marketing action, and was also the treatment in which *Timing* was most important in determining the relationship between past and future responses. In contrast, the No Mail control was the least effective action, and was also the condition in which the *Past Response* * *Timing* coefficient was smallest.[11]

**Summary**

We have proposed using the *Timing* of past response as an informative signal of dilution. The more slowly that responses have decreased over time, the less diluted the market. We then showed that the timing of past responses can inform the relationship between past and future responses. The information depends upon the effectiveness of the firm's future marketing action. Consider a highly effective marketing action. The extent to which a zip code is diluted is important for the relationship between the past responses and the future responses in that zip code. In contrast, if the marketing action is ineffective, then the effects of dilution on the relationship between past responses and future responses are less important. In our final set of analysis, we investigate whether the timing of past responses can improve the performance of targeting models.

## 5. Training Targeting Models Using the Timing of Past Responses

We focus on the problem of identifying which zip codes the retailer should send promotions to when prospecting for new customers.[12] The optimal policy depends upon predictions of how many households

---

[11] We note that if we focus on the *Past Response Rate* * *Timing* coefficient of the response rate models, the ordering of the coefficient across the three marketing actions is again consistent with the ordering of the average effectiveness of each marketing action in the Spring 2015 experiment. This ordering breaks in the Spring 2016 experiment, where the coefficients for the $25 paid promotion condition and the No Mail condition are not significantly different from zero (see Table 2). However, even in Spring 2016, it remains the case that the 120-day trial is the treatment in which *Timing* was most important in determining the relationship between past and future response rates.

[12] Ideally a firm would make household-level decisions, rather than zip code-level decisions. However, retailers do not have past purchasing histories for prospective customers (who have not previously purchased). This retailer uses demographic variables to make its targeting decisions. It purchases demographic measures aggregated to approximately 400 households (at the carrier route level). We then aggregate to the zip code level (approximately 13 carrier routes on average). The advantage of aggregating to the zip code level is that the response rates for each treatment are calculated more accurately in each zip code. The disadvantages of aggregation are less precise targeting, as well as the fact that the aggregation reduces the sample size (see Simester et al. 2019a for a discussion of these issues).

in a zip code will respond to each treatment. For example, if the firm is choosing between mailing a promotion (Mail) versus not mailing a promotion (No Mail), the targeting decision will be made on the basis of the difference in the number of predicted responses under these two actions:

$$s_3^{mail} - s_3^{no\ mail} = r\left(q^{mail} - q^{no\ mail}\right)(1 - q^{past})^2$$

As we have discussed (and illustrated), the timing of the past responses provides more information when the difference in the effectiveness of the marketing actions is larger (when $q^{mail} - q^{no\ mail}$ is larger).

In the previous section we reported the average response rate in the two experiments was highest for the 120-day trial condition, and lowest for the No Mail control.[13] Because the timing of past response is expected to be most informative when there is a larger difference in the effectiveness of the marketing actions, we expect that incorporating the *Timing* of past responses will contribute the largest performance improvement when choosing between the 120-day trial and the No Mail actions. In contrast, we expect the performance improvement to be smallest when the policy is choosing between the 120-day trial and the $25 paid actions. This comparison will form an important part of our analysis. We begin by describing a benchmark targeting model that does not use timing, and then describe how we modify this process to incorporate timing information.[14]

**Benchmark: No Timing Policy**

1. Use a sample of training data and OLS to separately estimate the following model for each marketing action:

   $$\textit{Response Rate}_z^{action} = \alpha + \beta_1\ \textit{Past Response Rate}_z + \gamma\ \textit{Demographics}_z + \varepsilon_{z,\cdot}$$

   As in Section 4, *Past Response Rate* is the response rate in the calendar year prior to the year of the experiment. The observations are weighted by the number of households in that zip code (that participated in the experiment).

2. Use the estimated coefficients from each model to predict the future response rate to each marketing action in each zip code in the implementation data.

3. Use these predicted response rates to calculate predicted responses, and therefore the predicted profit in each zip code under each marketing action. The profit calculation accounts for mailing costs and the average profit earned from each household that responds.

4. In each zip code in the implementation data, choose the action that has the highest predicted profit.

We modify this benchmark model by including the timing of past responses in Step 1.

---

[13] The difference in the effectiveness of each pair of marketing actions is reported in the Appendix.

[14] Before our research collaboration with this retailer, the retailer's standard approach to targeting prospective new customers was to estimate a propensity model. The firm did not train these propensity models using experimental data, and so did not segment customers using their responsiveness to the firm's marketing actions.

**Timing Policy**

1. Use the training data to separately estimate the following OLS model for each marketing action:

   $$\textit{Response Rate}_z^{action} = \alpha + \beta_1 \textit{ Past Response Rate}_z * \textit{Timing}_z + \beta_2 \textit{ Past Response Rate}_z + \beta_3 \textit{ Timing}_z + \gamma \textit{ Demographics}_z + \varepsilon_z$$

   *Past Response Rate* is defined above and *Timing* is the proportion of past responses in the prior calendar year that occurred in the second half of that prior year (see Section 4). The observations are again weighted by the number of households in that zip code (that participated in the experiment).

2. As above.
3. As above.
4. As above.

The only difference in the two procedures is that in the Timing Policy we include *Timing* and *Past Response Rate * Timing* as additional covariates in Step 1.[15]

**Estimators**

In Step 1 of the Timing and No Timing Policies, we estimate a separate OLS model for each marketing action. In practice, we are not just limited to using OLS; firms are exploring the use of increasingly sophisticated methods to perform this step. Helpfully, the firm that provided data for this study participated in a previous study investigating the effectiveness of a wide range of estimators.[16] The findings are documented in Simester at al. (2019a); Lasso consistently outperformed the other methods. For this reason, we will separately train models using OLS and Lasso. We next describe how we train the policies and compare their performance.

**Training and Evaluating the Policies**

For both models, we use the 412 zip codes in the Spring 2015 experiment as the training data and the 2,558 zip codes in the Spring 2016 experiment as the implementation data. Recall that in both experiments, there is a subsample of customers in each zip code that received each marketing action. We use the three Spring 2015 subsamples as zip code-level estimation samples to perform Step 1. We do this separately for the Timing Policy and the No Timing Policy using OLS, and then again using Lasso.

For each policy, these equations provide the expected response rates under each action in each zip code in the implementation data (this is Step 2). We use the predicted response rates to calculate the predicted profit in each zip code under each marketing action in Step 3, and then we choose the recommended action in Step 4. The predicted profits calculated in Step 3 use the actual mailing cost that

[15] Notice that the equation estimated in Step 1 of the No Timing Policy is Equation 4 in our earlier analysis, while the equation estimated in the Timing Policy is Equation 2.

[16] The methods included: Lasso, Finite Mixture Models, Hierarchical Clustering, Kernel Regression, k-NN, CHAID, SVM, XGBoost, Random Forest, and Neural Networks. The methods were trained using the same Spring 2015 training data that we use in this study.

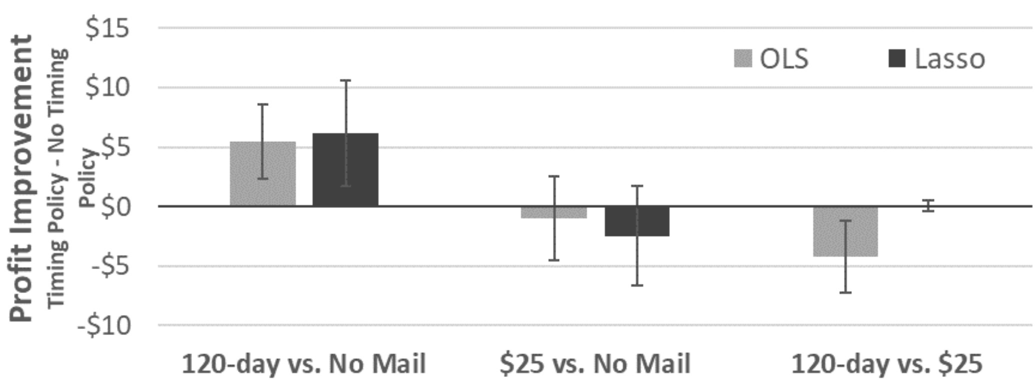
the firm incurred for each treatment and the average profits earned per response. When training the policies, these average profits were calculated using the actual Spring 2015 outcomes.

To evaluate the performance of a policy, we start by identifying the recommended action under the policy in each zip code in the implementation data (the Spring 2016 experiment). If a policy recommended the 120-day trial in zip code *z*, we calculate the actual number of responses received in zip code *z* among the households that received the 120-day trial in the Spring 2016 experiment. We use an analogous approach for the other two treatments. This allows us to estimate the number of responses that would have been received under each policy. We then estimate the profit that would have been earned from the policy by accounting for mailing costs and the average profit earned from each response.[17] We do this for both the Timing Policy and the No Timing Policy using OLS, and then again using Lasso.

**Performance Improvement**

In Figure 5, we compare the performance of the Timing and No Timing policies. In each case we construct three separate targeting policies, where each policy is restricted to choosing from only two actions: 120-day treatment vs. No Mail, $25 treatment vs. No Mail, and 120-day treatment vs. $25 treatment. Because the actual profits are confidential, we just report the *Profit Improvement*, calculated as the average profit per zip code in the Timing Policy minus the average profit per zip code in the No Timing Policy. In Figure 5 we report the findings for all 2,558 zip codes. In the Appendix we also report the findings for just the zip codes in which the two policies recommend different actions.[18]

**Figure 5. Performance Improvement**




The figure illustrates the *Performance Improvement*, calculated as the average profit per zip code in the Timing Policy minus the average profit per zip code in the No Timing Policy. Error bars indicate 95% confidence intervals. Additional details are provided in the Appendix.

---

[17] When training the policies, we calculate the profit per response using data from the Spring 2015 experiment. When evaluating the policies, we calculate the profit per response using either the Spring 2015 or Spring 2016 experiments. The results are almost identical (we report findings using the Spring 2016 data).

[18] For zip codes for which the two policies recommend the same actions, the difference in their performance is zero (Simester et al. 2019b). Focusing on the zip codes in which the policies recommend different actions isolates the differences in the two policies.

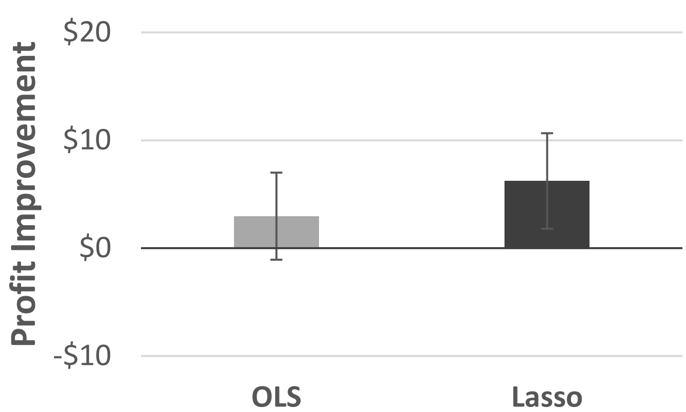
Unlike our earlier findings, the findings in Figure 5 represent an out-of-sample prediction. We train the policies using the Spring 2015 experimental data, and evaluate the resulting policies using the Spring 2016 data. The findings confirm that when choosing between the 120-day trial and No Mail, the timing information contributes a (very) significant improvement in the performance of the targeting models; $5.47 per zip code for OLS and $6.14 for Lasso, or equivalently 2.1 cents (OLS) and 2.5 cents (Lasso) increase for each household involved in the Spring 2016 experiment. For this retailer, a profit increase of 2.5 cents per prospective household represents a very substantial incremental profit each year (the retailer sends direct mail promotions to tens of millions of prospective customers each year).

When we focus on the other two-action decisions, $25 paid vs. No Mail, and 120-day trial vs. $25 paid, the Timing policy no longer outperforms the No Timing policy. Our earlier observations as well as the findings in the previous section help to explain why. Recall that we showed that the timing of past responses is particularly informative when the marketing actions are most effective. As a result, we expect *Timing* to be more valuable in training a targeting model when there is a larger difference in the effectiveness of a firm's marketing actions. The pattern of findings in Figure 5 is consistent with this result.

We emphasize that the larger profit improvement in the 120-day versus No Mail policies is not simply because it is easier to design a targeting policy when there is more variation in the effectiveness of the two actions. If this were the case, then this would benefit both the Timing and the No Timing policies. Instead, as we discussed and illustrated in the previous sections, the importance of timing (and dilution) depends upon how effective the firm's future marketing actions are. The large difference in the effectiveness of the 120-day and No Mail actions amplifies the importance of the timing information.

For completeness, we also trained policies that chose from all three marketing actions. These findings are reported in the Appendix and summarized in Figure 6. We see that when using Lasso, the timing information significantly improves the performance of the three-action targeting policy. There is also a performance improvement when using OLS, but the improvement is not statistically significant.

**Figure 6. Performance Improvement When Choosing from All Three Marketing Actions**




The figure illustrates the *Performance Improvement*, calculated as the average profit per zip code in the Timing Policy minus the average profit per zip code in the No Timing Policy. Error bars indicate 95% confidence intervals. Additional details are provided in the Appendix.

**Summary**

In the previous sections we showed that when zip codes are less diluted, then past response is a more positive signal of future response, particularly when marketing actions are effective. We also showed that the timing of past responses can provide a measure of dilution in a zip code. In this section, we combined these insights, and investigated whether the timing of past responses can improve the performance of a targeting model. We first illustrate how to include timing when training a targeting policy, and then confirm using two field experiments that timing can contribute significant performance improvements out of sample. The performance improvement depends upon the firm's action space; the improvement is largest when there is a large difference in the effectiveness of the marketing actions that the firm is choosing from.

## 6. Conclusions

Non-stationarity is a major obstacle to targeting in a batch-learning environment, where there is oftentimes a material delay between the collection of data and the implementation of policies based on it. While non-stationarity can arise for many reasons, we focus on the role of dilution. When customers can only adopt once, the pool of remaining customers will generally become more diluted over time, with fewer good prospects remaining. As a result, the response to the same marketing action will change over time. Unlike many other sources of non-stationarity, we show that dilution is a phenomenon that can be anticipated and measured. We also show that it can be an important factor that the literature has largely ignored.

The two key insights in the paper are (a) dilution may make high past response a less positive signal of the future response and (b) the timing of past responses may provide a measure of dilution. These insights are related; they focus on the relationship between responses in different time periods, and depend upon non-stationarity resulting from dilution. In the first insight, dilution allows us to make a forward-looking prediction. The second insight looks in reverse, by using the timing of past responses to measure of dilution.

We provide a simple illustrative model to explain why the timing of past responses can serve as a measure of dilution. The model also illustrates why we expect larger future responses from markets that are less diluted, and why dilution plays a more important role when the firm's marketing actions are more effective. When the marketing action is highly effective, then dilution has a larger impact on the number of responses received. However, when the marketing action is ineffective, then the effects of dilution are less important.

We show that this relationship can be identified empirically, not just when predicting future responses, but also when measuring how future responses vary with changes in marketing actions. Using data from two field experiments, we show that the timing of past responses can inform the relationship between past and future responses, and that the information depends upon the effectiveness of the firm's future marketing action.

The information about the timing of past responses can help to improve performance of targeting policies. We demonstrate that a targeting policy trained using the timing of past responses can significantly outperform a policy trained without using timing. The findings confirm that including the timing of past responses in targeting decisions can have a significant business impact, particularly when there is a larger difference in the effectiveness of the marketing actions.

While we use dilution to motivate the use of timing measures, we recognize that timing measures may also capture other sources of non-stationarity. This could include wearin or wearout of a promotion, together with dynamics due to the diffusion of word of mouth. As a result, the interpretation of the timing of past responses is complicated by multiple factors. However, irrespective of the source of this variation, we have shown that in a context where dilution matters, timing can improve predictions of future responses.

## References


Ascarza, Eva (2018), "Retention Futility: Targeting High-Risk Customers Might Be Ineffective," *Journal of Marketing Research* 55(1), 80–98.

Chae, Inyoung, Hernan A. Bruno, and Fred M. Feinberg (2019), "Wearout or Weariness? Measuring Potential Negative Consequences of Online Ad Volume and Placement on Website Visits," *Journal of Marketing Research*, 56(1) 57-75.

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

Fader, Peter S., and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12(3), 304-317.

Fader, Peter S., Bruce GS Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42(4), 415-430.

Farris, Paul W., Neil Bendle, Phillip E. Pfeifer, and David Reibstein (2010). *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*. Upper Saddle River, New Jersey: Pearson Education, Inc.

Goel, Sharad, and Daniel G. Goldstein (2013), "Predicting Individual Behavior with Social Networks, *Marketing Science*," 33(1), 82-93.

Gurvich, Itay, Mor Armony, and Constantinos Maglaras (2009), "Cross-selling in a Call Center with a Heterogeneous Customer Population," *Operations Research*, 57(2), 299-313.

Kumar, Vijaya (2008), "Customer Lifetime Value— The Path to Profitability," *Foundations and Trends in Marketing*, 2(1), 1-96.

Naik, Prasad A., Murali K. Mantrala, and Alan G. Sawyer (1998), "Planning Media Schedules in the Presence of dynamic Advertising Quality," *Marketing Science*, 17(3) 214-235.

Negoescu, Diana M., Kostas Bimpikis, Margaret L. Brandeau, and Dan A. Iancu (2018), "Dynamic Learning of Patient Response Types: An Application to Treating Chronic Diseases," *Management Science*, 64(8), 3469-3488.

Netzer, Oded, James M. Lattin, and Vikram Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27(2), 185-204.

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.

Rafieian, Omid, and Hema Yoganarasimhan (2020), "Targeting and Privacy in Mobile Advertising," *Working Paper*, available at SSRN 3163806.

Rossi, Peter E., Robert E. McCulloch, and Greg 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*, 36(4), 500-522.

Simester, Duncan I., Artem Timoshenko, and Spyros Zoumpoulis (2019a), "Targeting Prospective Customers: Robustness of Machine Learning Methods to Typical Data Challenges," *Management Science*, forthcoming.

Simester, Duncan I., Artem Timoshenko, and Spyros Zoumpoulis (2019b), "Efficiently Evaluating Targeting Policies: Improving Upon Champion vs. Challenger Experiments," *Management Science*, forthcoming.

Toubia, Olivier, Jacob Goldenberg, and Rosanna Garcia (2014), Improving Penetration Forecasts Using Social Interactions Data, *Management Science*, 60(12), 3049-3066.

U.S. Census (2019), http://www5.kingcounty.gov/sdc/FGDCDocs/ACS_B25001_HOUSINGUNITS_fgdc.htm. Retrieved on April 14, 2020.

Zhang, Jonathan Z. (2020), "Consumer Dynamics: Theories, Methods, and Emerging Directions," *Journal of the Academy of Marketing Science*, 1-31.

## Appendix

### Experiments: Additional Details

The paper uses data from three experiments, which were implemented by the same firm approximately six months apart:

1. Spring 2015 experiment
2. Fall 2015 experiment
3. Spring 2016 experiment

The experiments included many common features, but also included some important differences. We discuss the features of each experiment below. Data from the Spring 2015 and Fall 2015 experiments was previously used in two published papers (Simester et al. 2019a and 2019b), which focused on unrelated topics. Additional details of the experiment can be found in these papers.

**Spring 2015 Experiment**

The Spring 2015 experiment included six experimental conditions, of which we focus on three; a $25 discount off the regular price of a membership, a 120-day free trial member, and a No Mail control. The sample sizes were:

| | |
|---|---|
| 120-day trial: | 406,787 households |
| $25 paid: | 406,547 households |
| No Mail: | 415,895 households |

These households were all located in two regions of the country. Households were randomly selected to receive one of the six treatments. A comparison of the control variables confirms that the randomization was correctly implemented. The treatments were implemented on the front cover, inside front cover and back cover of a 48-page book of product-specific coupons. Each household received the same treatment twice; they received the first book of coupons in early February 2015, and then received an identical book of coupons in late March. We measured the outcome of the experiments using the number of responses received in the 63 days after the in-home date for the first mailing wave.

The Spring 2015 experiment is used in two places in the paper:

a. In Section 4 we use the data to study the relationship between past and future responses, the timing of past responses, and the effectiveness of the firm's marketing actions.

b. In Section 5 we use the data as training data when evaluating the two targeting models.

**Fall 2015 Experiment**

The Fall experiment included 11 experimental conditions, although we just use the three conditions that matched the Spring 2015 experimental treatments.

The households in the promotion treatments were mailed postcards that described the offer on both the front and back of the postcard. The postcards were sent in two waves 37 days apart. All of the households that received the postcard on the first wave also received the postcard on the second wave, and each household received the same version in both waves.

Prospective customers were grouped into carrier routes, which are groups of approximately 400 households located in close geographic proximity. They are literally the routes that individual letter carriers use to deliver mail. Each of the carrier routes is located within a single zip code and a typical zip code has approximately 13 carrier routes. The experimental treatments were randomly assigned to carrier routes. Every household in a carrier route received the same treatment as the other households in that carrier route, except if a household already had an active membership. Households that were active members before Wave 1 were removed from the mailing list (for both waves). A comparison of the control variables confirms that the randomization was correctly implemented.

Data from the Fall 2015 experiment was used in just one place in the paper. It provides data for Figure 2, where we show that the response in Spring 2016 varies systematically with the treatments that the households received in Fall 2015.

The USPS sometimes changes the carrier routes. When it does, it typically changes the carrier route identifier. As a result, not all of the carrier routes included in the Fall 2015 experiment had a matching identifier in the Spring 2016 experiment. In our analysis we restrict attention to carrier routes that had identical identifiers in both experiments. Below we report both the total number of carrier routes assigned to the four uniform treatments in Fall 2015, and the number of carrier routes used in our analysis in Figure 2.

**Number of Carrier Routes**

| | **$25 Paid** | **120-day free trial** | **No Mail Control** |
|---|---|---|---|
| Total in Fall Experiment | 1,026 | 1,003 | 1,062 |
| Used in Figure 2 | 745 | 806 | 729 |

**Spring 2016 Experiment**

To the best of our knowledge the data from the Spring 2016 experiment has not been used in any other academic paper. This experiment was designed to test how customers respond to different sequences of mailings. Households were randomly assigned to 19 conditions, in which the $25 paid condition, 120-day trial and No Mail Control were varied across three waves of mailings. The first and second waves were 21 days apart and the second and third waves were 42 days apart. The 19 experimental conditions are summarized in the table below.

| Condition | Wave 1 | Wave 2 | Wave 3 | Number of Households |
|---|---|---|---|---|
| 120_120_120 | 120 | 120 | 120 | 73,008 |
| 120_120_25 | 120 | 120 | 25 | 72,902 |
| 120_120_No | 120 | 120 | No | 73,240 |
| 120_25_120 | 120 | 25 | 120 | 73,249 |
| 120_25_25 | 120 | 25 | 25 | 73,152 |
| 120_25_No | 120 | 25 | No | 73,060 |
| 120_No_No | 120 | No | No | 73,506 |
| 25_120_120 | 25 | 120 | 120 | 73,646 |
| 25_120_25 | 25 | 120 | 25 | 73,710 |
| 25_120_No | 25 | 120 | No | 73,334 |
| 25_25_120 | 25 | 25 | 120 | 73,490 |
| 25_25_25 | 25 | 25 | 25 | 73,106 |
| 25_25_No | 25 | 25 | No | 73,288 |
| 25_No_No | 25 | No | No | 73,236 |
| No_120_No | No | 120 | No | 73,140 |
| No_25_No | No | 25 | No | 73,235 |
| No_No_120 | No | No | 120 | 73,136 |
| No_No_25 | No | No | 25 | 73,417 |
| No_No_No | No | No | No | 73,273 |

The assignment of households to conditions was independent of carrier routes. As a result, households in each carrier route received the same distribution of promotions (in expectation). A comparison of the control variables confirms that the randomization was again correctly implemented.

Households that were active members before Spring Wave 1 were again removed from the mailing lists. This provides the source of additional dilution between the Fall and Spring experiments.

The promotions were mailed using a single sheet of paper folded in half. The promotions were described on both sides of the outside of the folded paper. They were also described in a letter printed on the inside of the folded paper.

As we explain in Section 2, the households that participated in the Spring 2016 experiment were from the same carrier routes that participated in the Fall 2015 experiment. They were distributed across a much broader range of zip codes than the households that participated in the Spring 2015 experiment.

The Spring 2016 experiment is used in three places in the paper:

a. It provides data for Figure 2 in Section 2, where we show that the response in Spring 2016 varies systematically with the treatments that the households received in Fall 2015.

b. In Section 4 we use the data to study the relationship between past and future responses, the timing of past responses, and the effectiveness of the firm's marketing actions.

c. In Section 5 we use the data as the implementation data when evaluating the two targeting models.

In (a) we use data from all 19 experimental conditions. In (b) and (c) we just use data from Waves 1 and 2 in the following nine conditions:

| **Treatment** | **Experimental Condition** | **Wave 1** | **Wave 2** | **Wave 3** | **Number of Households** |
|---|---|---|---|---|---|
| 120-day trial | 120_120_120 | 120 | 120 | 120 | 73,008 |
| | 120_120_25 | 120 | 120 | 25 | 72,902 |
| | 120_120_No | 120 | 120 | No | 73,240 |
| $25 discount | 25_25_120 | 25 | 25 | 120 | 73,490 |
| | 25_25_25 | 25 | 25 | 25 | 73,106 |
| | 25_25_No | 25 | 25 | No | 73,288 |
| No Mail | No_No_120 | No | No | 120 | 73,136 |
| | No_No_25 | No | No | 25 | 73,417 |
| | No_No_No | No | No | No | 73,273 |

Notice that the three experimental conditions under each treatment received an identical treatment in Waves 1 and 2, but a different treatment in Wave 3. For this reason, the outcome measure we use when evaluating the two targeting policies restricts attention to the responses received in the 63 days between the in-home date for Wave 1 and the in-home date for Wave 3. This ensures that the findings are not confounded by the marketing actions implemented in Wave 3. When training the two targeting models using data from the Spring 2015 experiment, we also just use data from the first 63 days of that experiment (63 days from the in-home date of the first mailing wave).

## Control Variables

The control variables were all identified at the carrier route level (there is no variation across households within a carrier route). They included the following nine variables:

| | |
|---|---|
| Age | Age of head of household |
| Home Value | Estimated home value |
| Income | Estimated household income |
| Single Family | A binary flag indicating whether the home is a single-family home |
| Multi-Family | A binary flag indicating whether the home is a multi-family home |
| Distance | Distance to nearest store for this retailer |
| Comp. Distance | Distance to nearest competitors' store |
| F Flag | Binary flag indicating whether the retailer considers the zip code "far" from its closest store |
| M Flag | Binary flag indicating whether the retailer considers the zip code a "medium" distance from its closest store |

**Summary Statistics for the Control Variables**

| | Average | Std. Error |
|---|---|---|
| Age | 57.03 | 0.09 |
| Home Value | 255,259 | 3,325 |
| Income | 86,848 | 1,199 |
| Single Family | 78.6% | 0.5% |
| Multi-Family | 21.0% | 0.5% |
| Distance | 10.30 | 0.17 |
| Comp. Distance | 11.68 | 0.25 |
| F Flag | 55.56% | 0.92% |
| M Flag | 29.69% | 0.74% |

The table reports the average and standard error for each of the control variables. The unit of analysis is a zip code and the sample size is 2,558 zip codes.

## The Information that Past Responses Provide About Future Responses

## Customer Acquisition File

We use the customer acquisition file to investigate the relationship between responses in the 2016 calendar year and responses in the previous two years (2014 and 2015). We show how this relationship depends upon the timing of past responses. We first describe how we construct measures of past responses and timing in this setting:

| | |
|---|---|
| $Response_z$ | Number of new customers acquired in 2016 in zip code *z*. |
| $Past\ Response_z$ | Number of new customers acquired in both 2014 and 2015 in zip code *z.* |
| $Timing_z$ | The proportion of *Past Response* acquired in 2015 (in zip code *z*). |

We also constructed two response rate measures:

| | |
|---|---|
| $Response\ Rate_z$ | The *Response* in zip code *z* divided by the number of households in that zip code.[19] |
| $Past\ Response\ Rate_z$ | The *Past Response* in zip code *z* divided by the number of households in that zip code. |

We use these measures to estimate the following zip code-level OLS models (focusing on either responses or response rates):

**Responses**

$$Response\ 2016_z = \alpha + \beta_1\ Past\ Response_z * Timing_z + \beta_2\ Past\ Response_z + \beta_3\ Timing_z + \varepsilon_z \quad (A1)$$

**Response Rates**

$$Response\ Rate\ 2016_z = \alpha + \beta_1\ Past\ Response\ Rate_z * Timing_z + \beta_2\ Past\ Response\ Rate_z + \beta_3\ Timing_z + \varepsilon_z \quad (A2)$$

The coefficient of interest in both models is $\beta_1$, which measures how the relationship between past and future responses varies with *Timing*. If the *Timing* ratio is large, indicating less dilution, we expect that past responses will be a stronger signal of future response. This interpretation argues that this interaction coefficient should be positive. The estimation sample in each model includes 11,752 zip codes, and we weight the response rate model using the number of households in each zip code.[20] The coefficients are reported in the table below.

[19] The retailer did not provide demographic information for all of these zip codes. Therefore, for the number of households we use an estimate provided by the U.S. census. It estimates the number of households in a zip code between 2013 and 2017. For more details see U.S. Census, 2019.

[20] From the 13,001 zip codes included in the analysis in Figure 2, we exclude 1,237 zip codes for which the number of housing units is missing, and 12 zip codes for which the number of housing units is less than the number of

**Timing and the Relationship Between Past and Future Responses**

| | Responses | Response Rates |
|---|---|---|
| *Past Response (Rate) * Timing* | 0.4403** | 0.4501** |
| | (0.0167) | (0.0176) |
| *Past Response (Rate)* | 0.2488** | 0.2419** |
| | (0.0083) | (0.0087) |
| *Timing* | 36.4317** | 0.0047** |
| | (9.6027) | (0.0009) |
| *Intercept* | -10.6142* | -0.0013** |
| | (4.5770) | (0.0004) |
| $R^2$ | 0.8998 | 0.8795 |

The table reports the coefficients from estimating Equations A1 and A2. The unit of analysis is a zip code and the sample size in both models is 11,752 zip codes. In the response rate model the observations are weighted by the number of housing units in that zip code. Standard errors are in parentheses. **Indicates significantly different from zero ($p < 0.01$). *Indicates significantly different from zero ($p < 0.05$). †Indicates significantly different from zero ($p < 0.10$).

The coefficient of interest ($\beta_1$) is positive and significant in both models. This is consistent with our interpretation that past responses are a stronger indicator of future responses when there is less evidence of dilution. In contrast, when there is a lot of dilution, indicated by a larger decrease in past responses over time, then past responses become a weaker signal of future responses.

As a robustness check, we repeated this analysis using different periods to measure timing, including half-year periods. We observed the same pattern of results. Furthermore, in the response rate model we adjusted the response rate in the dependent variable to remove households that responded in prior years from the denominator. The coefficient of interest was again positive and significant.

---

responses in 2016. Many of these discrepancies are likely due to errors (by the census) in the estimation of the number of households in a zip code. The US Postal Service also changes the design of zip codes over time.

**Timing and the Relationship Between Past and Future Responses: Median Splits**

| | Responses | | Response Rates | |
|---|---|---|---|---|
| | Spring 2015 | Spring 2016 | Spring 2015 | Spring 2016 |
| **Below Median** | | | | |
| *Past Response (Rate)* | 0.0695** | 0.0015** | 0.0859** | 0.1104** |
| | (0.0021) | (0.0001) | (0.0069) | (0.0077) |
| $R^2$ | 0.9181 | 0.2361 | 0.8179 | 0.3684 |
| Sample Size | 206 | 1,279 | 206 | 1,279 |
| **Above Median** | | | | |
| *Past Response (Rate)* | 0.0829** | 0.0027** | 0.1378** | 0.1496** |
| | (0.0025) | (0.0002) | (0.0089) | (0.0085) |
| $R^2$ | 0.9223 | 0.2763 | 0.8477 | 0.3954 |
| Sample Size | 206 | 1,279 | 206 | 1,279 |

The table reports the coefficients ($\beta_1$) from estimating Equations 3 and 4 separately using median splits of the *Timing* measure. Each coefficient is obtained from a separate model. Error bars indicate 95% confidence intervals.

**The Effectiveness of the Firm's Marketing Actions: $R^2$ Values**

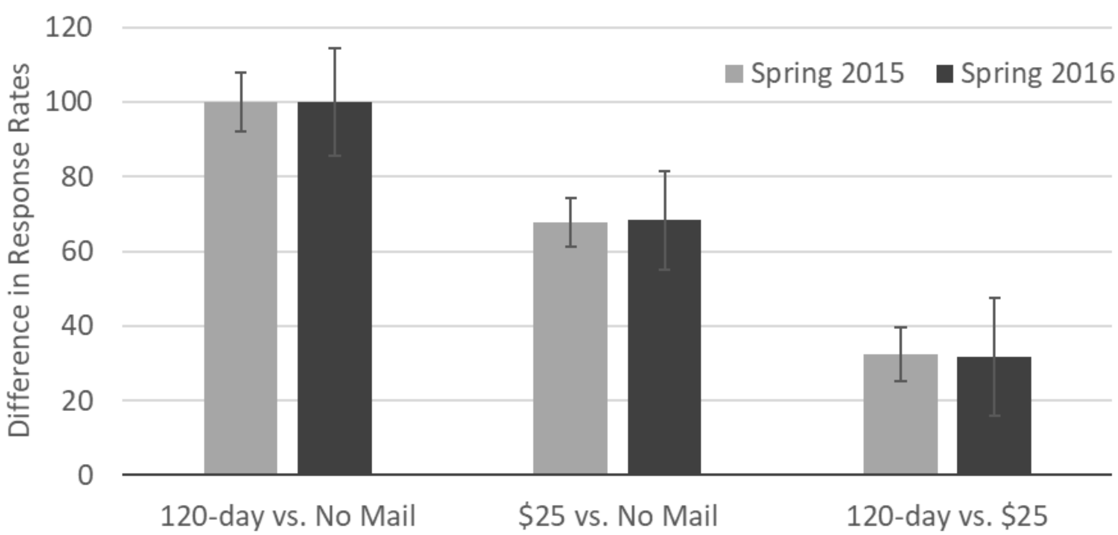
| | Responses | | Response Rates | |
|---|---|---|---|---|
| Experimental Condition | Spring 2015 | Spring 2016 | Spring 2015 | Spring 2016 |
| 120-day Trial | 0.8816 | 0.1710 | 0.7349 | 0.2192 |
| $25 Paid | 0.8572 | 0.1841 | 0.6919 | 0.2164 |
| No Mail | 0.7585 | 0.1230 | 0.5719 | 0.1300 |

The table reports the $R^2$ values from estimating Equations 1 and 2. The estimated *Past Response (Rate) * Timing* coefficients are reported in Table 2. The unit of analysis is an experiment (Spring 2015 or Spring 2016) x zip code x marketing action (120-day, $25, or No Mail). The sample sizes are 412 (Spring 2015) and 2,558 (Spring 2016). The observations in the response rate models are weighted by the number of households in the zip code (that participated in the experiment).

**The Difference in the Effectiveness of Each Pair of Marketing Actions**

| | Difference in Response Rates | |
|---|---|---|
| | **Spring 2015** | **Spring 2016** |
| 120-day vs. No Mail | 100.00 | 100.00 |
| | (3.96) | (7.39) |
| $25 vs. No Mail | 67.63 | 68.34 |
| | (3.35) | (6.77) |
| 120-day vs. $25 | 32.37 | 31.66 |
| | (3.64) | (8.07) |
| Number of zip codes | 412 | 2,558 |

The table reports the difference in the response rates for pairs of marketing actions in the Spring 2015 and Spring 2016 experiments (averaged across zip codes). These differences are indexed to equal 100 in the 120-day vs. No Mail comparison. Standard errors are in parentheses.




The figure illustrates the difference in the response rates for pairs of marketing actions in the Spring 2015 and Spring 2016 experiments, which is also reported in the table above. The differences are first calculated at the zip code level, and then averaged across the 412 Spring 2015 and 2,558 Spring 2016 zip codes. The differences are then indexed to equal 100 in the 120-day vs. No Mail comparison. Error bars indicate 95% confidence intervals.

**Average *Profit Improvement***

| | | *Profit Improvement* | | |
|---|---|---|---|---|
| | | **All 2,558 Zip Codes** | **Zip Codes where Policies Differ** | **Nbr. of Zip Codes where Policies Differ** |
| **OLS** | **Two Actions** | | | |
| | 120-day Trial or No Mail | \$5.47** ($1.60) | \$152.62** ($41.18) | 95 |
| | \$25 Paid or No Mail | -\$1.03 ($1.80) | -\$24.67 ($39.41) | 128 |
| | 120-day Trial or \$25 Paid | -\$4.17 ($1.54) | -\$192.61 ($49.30) | 101 |
| | **Three Actions** | | | |
| | 120-day Trial, \$25 Paid, or No Mail | \$2.96 ($2.06) | \$61.01† ($35.63) | 174 |
| **Lasso** | **Two Actions** | | | |
| | 120-day Trial or No Mail | \$6.14** ($2.25) | \$148.81** ($56.99) | 91 |
| | \$25 Paid or No Mail | -\$2.49 ($2.13) | -\$80.80 ($64.60) | 90 |
| | 120-day Trial or \$25 Paid | \$0.08 ($0.22) | \$95.02 ($259.14) | 3 |
| | **Three Actions** | | | |
| | 120-day Trial, \$25 Paid, or No Mail | \$6.23** ($2.26) | \$147.68** ($55.74) | 94 |

The table reports the *Profit Improvement*, calculated as the average profit per zip code in the Timing Policy minus the average profit per zip code in the No Timing Policy. **Indicates there is a profit improvement that is significantly different from zero ($p < 0.01$). *Indicates there is a profit improvement that is significantly different from zero ($p < 0.05$). †Indicates there is a profit improvement significantly different from zero ($p < 0.10$).