Can Non-tiered Customer Loyalty Programs Be Profitable?

Arun Gopalakrishnan

Zhenling Jiang

Yulia Nevskaya

Raphael Thomadsen¹

Washington University in St. Louis

¹ Corresponding Author. Olin Business School, Washington University in St. Louis, One Brookings Dr., Knight Hall 404, St. Louis, MO 63132. thomadsen@wustl.edu. 314-935-3573.
Can Non-tiered Customer Loyalty Programs Be Profitable?

Abstract

In this paper, we study the impact of launching a non-tiered customer loyalty program on customer behavior. Specifically, we estimate how much a non-tiered loyalty program affected consumers’ spending per visit, frequency of visits, and attrition rates, as well as the overall customer value. Our estimation is aided by three aspects of our data that help us avoid selection bias in our estimates: (1) We track a cohort of customers before and after program introduction; (2) we exploit the firm’s exogenous assignment of some customers to the program; and (3) we develop a model based on a Hidden Markov Model that allows for appropriate accounting of program effects on frequency and attrition. We find the program introduction results in an almost 19% gain in customer value over a five-year horizon, an effect that is considerably larger than what has been previously found in the literature for non-tiered loyalty programs. We suggest our effect size is larger because of our ability to measure the impact of the program in preventing attrition among program members: We show that the program’s reduction in attrition leads to a 14% increase in customer value, accounting for over 70% of the program’s total lift. We find that overall program effects on frequency are smaller, with increased frequency leading to only a 4% increase in customer value. The impact of the loyalty program on spending per visit is negligible. Further, we find the program effects are heterogeneous across consumers of varying pre-program involvement. The relative lift in customer value is highest for the low-and high-frequency segments and much smaller for the moderate-frequency segment. This finding is driven by a similar pattern in the effect of the program on attrition. Frequency effects, however, are higher for customers who were already high-frequency visitors compared to lower frequency prior to program introduction.
1. Introduction

Customer loyalty programs are used by a wide range of businesses. In 2013, the average US household belonged to 21.9 loyalty programs and actively participated in 9.5 of them (Berry, 2013). An important attribute of a loyalty program is whether it provides a tiered (i.e., increased rewards for reaching higher thresholds) or non-tiered reward structure. Hotel chains and airlines typically offer tiered programs, in which mechanisms such as economic lock-in due to increasing benefits and consumer self-signaling (e.g., Drèze and Nunes, 2009; Orhun and Guo, 2019) may be at play. By contrast, non-tiered programs, such as “buy 10 get 1 free,” or “$X off for every $Y of spend” are popularly used in retail and service industries such as grocery stores, coffee shops, sandwich shops, and golf courses. They are also found in credit-card reward programs. Unlike a tiered program, it is less obvious how a non-tiered program may increase customer demand or be profitable because status cannot be earned. Further, a non-tiered program provides much weaker economic lock-in from skipping a visit to the particular business, because although the customer receives no credit for the one skipped visit, she incurs no loss of the increasing value for future visits. However, consumers may respond to rewards programs for psychological reasons. For example, the presence of the rewards program can make the customer feel emotionally connected to a particular firm, which leads to the customer visiting the firm more often.

The existing literature has typically found small or statistically insignificant effects from non-tiered loyalty programs. For example, Sharp and Sharp (1997) find only a weak loyalty-program impact through repeat purchases in retail outlets, and the effect is not consistently observed for all brands. Hartmann and Viard (2008) find that a “buy 10 get 1 free” loyalty program does not create significant switching costs for members. Lewis (2004) finds a 2% revenue increase from a frequency loyalty program. Leenheer et al. (2007) find the non-tiered loyalty program they study increases a store’s share of wallet by about 4%.²

---

² Taylor and Neslin (2005) also find that a similar program in which shoppers who spent a certain amount of money received a turkey, increased sales by approximately 6% during the program’s eight-week period, and by 1.8% over the seven weeks after the promotional period. These benefits have to be balanced against the cost of the program; although the paper does not state the costs of the program, a reasonable estimate of costs might be in the range of 2% – 4% based on the average spending reported in the paper and the approximate cost of a turkey, which would put the
One difficulty in measuring the impact of a loyalty program is that each customer cannot be simultaneously observed while in and out of the loyalty program. Measuring the program effect on any behavioral aspect (frequency, attrition, or spending) is then contingent on the construction of a plausible counterfactual. The issue would be easily resolved if customers could be randomly assigned to a treatment (“join the program”) and a control (“do not join the program”) group. However, in practice, customers self-select once a loyalty program is launched, leading to large selection effects (van Heerde and Bijmolt, 2005; Leenheer et al., 2007). Studies that do not fully account for selection effects (e.g., Bolton et al., 2000; Lal and Bell, 2003) have shown sizeable effects, which are likely to be upwardly biased. Leenheer et al. (2007) address such selection issues by using instrumental variables, but such a solution depends on having strong and valid instruments, which are hard to find. Other papers consider only the behavior of loyalty-program members (Lewis, 2004; Liu, 2007; Hartmann and Viard, 2008; Kopalle et al., 2012; Stourn et al., 2015), and use a model based on rational economic behavior to identify the program’s impact. As such, they capture the modeled effect of the loyalty program, but may not capture the entire impact, because a consumer’s responses to a loyalty program can go beyond those based on pure “economic utility” (Henderson et al., 2011). For example, the primary economic rationale for a non-tiered loyalty program would be reward redemption, which then suggests that high-frequency customers would mainly benefit, whereas low-frequency customers would see less value in being part of the program. However, other psychological benefits may accrue simply from being a member of a loyalty program even if one does not qualify for its rewards. We wish to allow for this possibility in our analysis.

The goal of this paper is to measure the effectiveness of a non-tiered loyalty program using data from a men’s hair-salon chain that introduced such a program in the time interval during which the data were collected. The overall program effectiveness is measured by the change in the time-discounted customer value over a five-year horizon that is due to the loyalty program. We consider how the program influences the three aspects of consumer behavior: spending per visit, visit frequency, and attrition rate. 

benefit in a range similar to that of the other papers listed here. The limited time nature of these programs may also affect the size of the measured impact.
Furthermore, we seek to understand the heterogeneous impact of the loyalty program on different types of customers. The main contribution of this paper is to show that our unique dataset, combined with a careful modeling approach to account for selection biases, shows a much larger (18.9%) increase in customer value from a non-tiered loyalty program than has previously been found in the literature. Put differently, this estimate suggests that, on average, a customer who becomes a loyalty-program member would contribute 18.9% more to the firm’s discounted five-year revenues than the same customer would contribute if she were not a member of the loyalty program. A large portion of this impact is driven by reducing the attrition rate, which has received little attention in the literature.

We utilize a number of novel aspects in our data and estimation approach. First, we collect data for a cohort of over 5,500 customers who were acquired by the firm around the same time before the program introduction, which eliminates unneeded variation among customers on the length of history with the company. All subsequent-visit data from these consumers are captured for a period of 30 months, during which time the loyalty program was introduced. Having consumer behavior before and after the program’s introduction allows us to control for consumer heterogeneity and selection effects. Second, we exploit an institutional detail in our setting—some customers were automatically enrolled in the loyalty program at the program introduction, such that their timing of joining the program is exogenous. We further divide customers into segments based on their pre-program visit frequency, and use this metric to create matched segments of automatically enrolled customers (which we hereafter refer to as “automatic members”) and non-enrolled customers\(^3\) (which we refer to as “non-members”). Finally, we bring together the evaluation of non-tiered loyalty programs and the literature on customer-lifetime-value models (e.g., Schmittlein et al., 1987; Fader et al., 2010; Ascarza and Hardie, 2013). Specifically, we develop a modeling framework based on a Hidden Markov Model (HMM) (e.g., Netzer et al. 2008) that integrates customer attrition and frequency of visits, which allows for the estimation of the program’s overall effectiveness. By estimating

---

\(^3\) Although non-members’ decision not to join the loyalty program could be strategic, we show that matching automatic members and non-members on visit frequency prior to the program results in very similar pre-program behaviors. The descriptive analysis in section 2 also reveals frequency results similar in magnitude to our main results once we control for consumer-level fixed effects, which should mitigate much of the selection effect.
separate HMM-based models for each of these segments (based on the number of pre-program visits) and
customer types (automatic- versus non-members), we allow for a flexible accounting of heterogeneity by
avoiding “pooling” data from infrequent and frequent customers. We compute the change in behavior (e.g.,
attrition and frequency) before and after the program introduction for automatic and non-members, and take
the difference in the change between the automatic and non-members. In this sense, our approach is akin to
the difference-in-differences (DID) estimator that is increasingly popular in the marketing literature (e.g.,
Ma et al., 2013), but with the difference that the average behaviors in this paper are captured by functions
of model parameters rather than directly from averages from the data.

We find the loyalty program has a minimal impact on spending per visit, and increases visit
frequency by less than 4%. The largest effect of the program is in attrition prevention, that is, preventing
customers from entering into a long hiatus from patronizing the hair salon while getting haircuts elsewhere.
We find that the loyalty program reduces the attrition rate by 22%. Together, the overall effect of the
program on customer value over a five-year horizon is 19%, which can be decomposed into approximately
14% coming from reduced attrition and 4% coming from increased frequency (with the rest of the effect
coming from the complementarities between the two). Our results also reveal heterogeneous treatment
effects of the program in terms of “treatment on the treated.” Attrition prevention is largest for very-low-
or very-high-frequency customers (classified using pre-program data) rather than moderate-frequency
customers. The program’s effect on visit frequency is, for the most part, higher for high-frequency
customers. In total, these effects lead to a relative lift in customer value that is highest for the low- and
high-frequency segments and much smaller for the moderate-frequency segment.

The rest of the paper is organized as follows. In section 2, we describe our data, the structure of the
loyalty program, and some descriptive analyses. In section 3, we describe the model specification of the
HMM-based approach we use for analysis. In section 4, we discuss the estimated model and analyze the
impact of the loyalty program on customer value. We present our conclusions in section 5.
2. Dataset and Descriptive Analysis

In this section, we first describe our data in section 2.1. We then provide some descriptive analyses in section 2.2 that demonstrate the impact of the loyalty program as a precursor to a HMM-based model that we introduce later in section 3.

2.1 Data Set

The empirical analysis is based on a dataset obtained from a chain of men’s hair salons. We observe a cohort of customers acquired between six and nine months before the launch of a loyalty program. All visits from these customers are captured for a period of 30 months. In the 10th month (of the 30-month period), the company introduced a customer loyalty program. Therefore, we have observations of customers’ visit behavior both before and after the launch of the loyalty program. Each transaction record contains a unique customer identification number, the date of the visit, the dollar amount spent, the services and products purchased, and any applied discounts.

Members of the loyalty program receive a $5-off reward coupon via email for every $100 they spend (across visits) on hair services and hair-care products. When the customers want to redeem the coupons, they are required to bring the coupon to the store or show the coupon email on a cellphone. Given that an average transaction value is around $21, customers typically need to visit five times to earn a coupon. The loyalty program does not have a membership fee. To become a member, the customer needs to provide an email address. Once in the program, members start to accrue spending toward receiving a reward coupon.

In our dataset, the price of a haircut increased by $1 a month after the loyalty program was launched, which is an approximately 5% increase. To normalize the analysis before and after the price increase, we added this price increase to the amounts spent before the price increase to avoid an erroneous inference that the loyalty program may have increased spending if it had not been accompanied by an increased demand

---

4 These customers do not have a visit during the six-month window prior to their observed first visit in our dataset. Hence, we assume these customers are newly acquired during their first observed visit.
5 Some customers may fail to redeem available coupons, due to forgetting or taking another discount that cannot be combined with the coupon.
for services (or products). Based on economic theory, we expect the price change to reduce the frequency of visits and the retention rate. Thus, our finding of the loyalty program’s positive effect on the visit frequency and retention should be viewed as a lower bound on its impact. Alternatively, one can note that because the introduction of a loyalty program generally pairs a new program with the introduction of the discount from that program, assessing whether the success of a loyalty program is due to the presence of the program or the corresponding price discount is hard. In our case, any effect we find must come from the presence of the program itself, because the total price to the customers in the post-program period is always at least as high as it is in the pre-program period.

Our empirical analysis leverages a unique group of loyalty-program members, whom we call automatic members. These customers provided their email addresses and agreed to receive marketing messages before the loyalty-program introduction (and without awareness that a loyalty program would be introduced). When the loyalty program was introduced, the firm signed them up automatically for the program; therefore, the timing of joining the program is exogenous for these customers. In many settings, customers can join a loyalty program whenever such a program is available. For that reason, an analyst should be concerned about the possibility of a dynamic selection bias: Customers may select the timing of when to join the program in response to a shift in their demand for haircut services. Due to this self-selection, being a member might coincide with higher demand for services, but the program would not cause the demand increase. Using automatic members as the treatment group minimizes these dynamic selection-bias concerns. Of course, other types of selection bias can remain in our data even after using automatic members, and we discuss how we handle these other types of selection biases when we discuss the estimation.

This cohort consists of 5,544 customers. Panel A of Table 1 shows the composition of the customer cohort. Approximately 38% of customers are members of the loyalty program, most of whom are automatic members who were enrolled by the firm at program launch. About 2% of customers (or 5% of members)
redeemed a reward coupon during our data period. Note that only these redeeming members entail costs to the company from running the loyalty program.⁶

### Table 1: Summary Statistics Customer Cohort Composition

#### Panel A: Customer Cohort Composition

<table>
<thead>
<tr>
<th></th>
<th>Number of</th>
<th>Share of</th>
<th>Average Number</th>
<th>Share of All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Customers</td>
<td>Customer Base</td>
<td>Transactions</td>
<td>Transactions</td>
</tr>
<tr>
<td>All Customers</td>
<td>5,544</td>
<td>100%</td>
<td>6.4</td>
<td>100%</td>
</tr>
<tr>
<td>Non-members</td>
<td>3,413</td>
<td>62%</td>
<td>4.9</td>
<td>47%</td>
</tr>
<tr>
<td>Members</td>
<td>2,131</td>
<td>38%</td>
<td>8.8</td>
<td>53%</td>
</tr>
<tr>
<td>Automatic Members</td>
<td>1,769</td>
<td>32%</td>
<td>7.4</td>
<td>37%</td>
</tr>
<tr>
<td>Redeeming Members</td>
<td>116</td>
<td>2%</td>
<td>23.6</td>
<td>8%</td>
</tr>
</tbody>
</table>

#### Panel B: Summary Statistics of Key Variables of Interest

<table>
<thead>
<tr>
<th></th>
<th>Net Spending per Visit ($)</th>
<th>Gross Spending per Visit ($)</th>
<th>Days between Visits</th>
<th>6-Month Attrition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Customers</td>
<td>20.63</td>
<td>20.66</td>
<td>50.14</td>
<td>17.83%</td>
</tr>
<tr>
<td>Non-members</td>
<td>20.18</td>
<td>20.19</td>
<td>54.44</td>
<td>24.35%</td>
</tr>
<tr>
<td>Members</td>
<td>21.03</td>
<td>21.09</td>
<td>46.83</td>
<td>12.02%</td>
</tr>
<tr>
<td>Automatic Members</td>
<td>21.11</td>
<td>21.16</td>
<td>47.21</td>
<td>15.09%</td>
</tr>
<tr>
<td>Redeeming Members</td>
<td>21.82</td>
<td>22.21</td>
<td>34.70</td>
<td>1.32%</td>
</tr>
</tbody>
</table>

Panel B of Table 1 shows the mean of the key variables of interest: net spending per visit (which takes into account any applied discounts), gross spending per visit, days between visits, and the attrition rate, which we define for the purposes of the descriptive analysis as an event where a customer does not visit for over 182 days. Note that we do not impose any time limit for attrition in our main analysis—this assumption is purely created to allow us to give some descriptive sense of the data. Compared to non-members, automatic members spend more, visit more often, and are less likely to have a hiatus from visits for six months or longer. We note, of course, that these summary statistics span both pre- and post-program introduction periods and therefore may reflect both the treatment effect of the loyalty program and the pre-program differences between automatic members and non-members.

⁶ The program’s costs are all fixed, except for the redemption costs.
2.2 Descriptive Analysis

In this subsection, we present some descriptive analysis about how the loyalty program affects the spending, visit frequency, and attrition of customers. The purpose of this analysis is two-fold. First, we seek to demonstrate that the patterns of behavior changes are found in the underlying data and are not simply a result of the assumed model. Second, our modified HMM framework focuses on changes in frequency and attrition, and not on changes in spending. We show here that the loyalty program has neither a statistically nor economically significant impact on spending. This insignificant impact on spending is likely a function of the industry we study, men’s haircuts, where most customers just want the basic service.

Spending per Visit

To analyze the impact of the program on spending, we compare both net and gross spending for automatic members before and after the introduction of the loyalty program, using the spending per visit of non-members over time as a non-parametric control for time trends. We do so by estimating a DID measure of how spending per visit changes after program introduction for automatic members compared to non-members.

We execute this DID using two separate strategies to account for possible selection effects (i.e., automatic members may behave differently from non-members even before the program introduction). The first strategy is to use customer fixed effects. We run a regression on how spending per visit changes for automatic members compared to non-members after the program introduction. Specifically, we run the following regression:

\[ Y_{ik} = \beta_{ap} \cdot M_i \cdot AP_{ik} + \beta_{w} \cdot W_k + \alpha_i + \epsilon_{ik}, \]

where \( Y_{ik} \) is customer \( i \)'s spending on visit \( k \), \( M_i \) is an indicator that equals 1 if customer \( i \) is an automatic member and 0 if non-member, \( AP_{ik} \) is an indicator that equals 1 if customer \( i \) visit to \( k \) occurs after the program introduction, \( W_k \) is a vector of year-week dummies that represent a common time trend (and seasonality) between automatic and non-members, \( \alpha_i \) is a customer fixed effect, and \( \epsilon_{ik} \) is an error term.
The main parameter of interest, $\beta_{ap}$, is a DID-type measure that is identified by how much the spending by automatic members changes after the program introduction compared to spending shifts of non-members.

The second strategy we use is to match automatic and non-members by their number of salon visits before the program is introduced. We divide the customers into seven matched segments—automatic members and non-members who have one, two, three, four, five, six, or seven or more visits before the program introduction. We also use this approach to account for customer heterogeneity in our model (section 3). The number of customers in each segment is shown in Table 2. Note that we do not include the 362 members who self-selected into the loyalty program (as opposed to being automatically enrolled by the firm) in our descriptive or model-based analysis. We also drop the 321 customers who made seven or more trips to the firm before the loyalty program was implemented, because these customers vary widely in the number of visits they made pre-program, and we have too few customers with any specific number of visits to conduct our full model analysis on those customers and get reliable estimates. The descriptive results are very similar if we include these customers, but we wish to be consistent regarding customers we include in the descriptive and the model analysis.

<table>
<thead>
<tr>
<th>Pre-program Visit Frequency</th>
<th>Automatic Members</th>
<th>Non-members</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>685</td>
<td>1763</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>605</td>
</tr>
<tr>
<td>3</td>
<td>237</td>
<td>355</td>
</tr>
<tr>
<td>4</td>
<td>173</td>
<td>267</td>
</tr>
<tr>
<td>5</td>
<td>126</td>
<td>163</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>107</td>
</tr>
<tr>
<td>7+</td>
<td>168</td>
<td>153</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,769</strong></td>
<td><strong>3,413</strong></td>
</tr>
</tbody>
</table>

The specific regression we run is as follows:

$$Y_{itk} = \beta_{ap} \cdot M_t \cdot AP_{tk} + \beta_m \cdot M_t + \beta_w \cdot W_{tk} + s_t + \epsilon_{tk}, \quad (2)$$
where $s_i$ is the segment fixed effect (based on the number of pre-program visits). The other variables are defined as above. In this specification, the segment fixed effects capture the customer heterogeneity. In addition, $\beta_m$ also captures differences between automatic members and non-members.

Table 3. Spending per Visit

<table>
<thead>
<tr>
<th></th>
<th>Net Spending Per Visit</th>
<th>Gross Spending Per Visit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Program Start*Automatic Members</td>
<td>-0.1234</td>
<td>-0.0771</td>
</tr>
<tr>
<td>($\beta_{ap}$)</td>
<td>(0.1705)</td>
<td>(0.1161)</td>
</tr>
<tr>
<td>Automatic Members</td>
<td>1.0059***</td>
<td>1.0046***</td>
</tr>
<tr>
<td></td>
<td>(0.2133)</td>
<td>(0.2138)</td>
</tr>
<tr>
<td>Number of Visits: 2</td>
<td>-0.1476***</td>
<td>-0.1502***</td>
</tr>
<tr>
<td></td>
<td>(0.0322)</td>
<td>(0.0318)</td>
</tr>
<tr>
<td>Number of Visits: 3</td>
<td>-0.1885***</td>
<td>-0.1897***</td>
</tr>
<tr>
<td></td>
<td>(0.0378)</td>
<td>(0.0379)</td>
</tr>
<tr>
<td>Number of Visits: 4</td>
<td>-0.0119</td>
<td>-0.0108</td>
</tr>
<tr>
<td></td>
<td>(0.0410)</td>
<td>(0.0408)</td>
</tr>
<tr>
<td>Number of Visits: 5</td>
<td>0.3154***</td>
<td>0.3317***</td>
</tr>
<tr>
<td></td>
<td>(0.0393)</td>
<td>(0.0389)</td>
</tr>
<tr>
<td>Number of Visits: 6</td>
<td>1.0503***</td>
<td>1.0802***</td>
</tr>
<tr>
<td></td>
<td>(0.0446)</td>
<td>(0.0445)</td>
</tr>
<tr>
<td>Individual Dummies</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Year+Week Dummies</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>22,231</td>
<td>22,231</td>
</tr>
</tbody>
</table>

Note: Cluster-Robust standard errors

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 3 shows the estimation results. Columns 1 and 2 show the impact of the program on net spending per visit. Column 1 shows the results from equation (1) with individual fixed effects, and column 2 uses the segment-level fixed effects from equation (2). The program’s impact on net spending is not statistically or economically significant. Columns 3 and 4 show the same analysis for gross spending per visit, which
allows us to examine if the null effect for net spending is due to the reward coupon or a lack of change in the services that are purchased. The coefficient $\beta_{\text{ap}}$ remains statistically and economically insignificant, indicating the loyalty program has a negligible effect on spending per visit. Further, the segment-level fixed effects show only highly frequent visitors (with six or more visits in the pre-program period) spend more than the least frequent visitors.

*Visit Frequency*

We next assess the impact of the program on visit frequency, which we represent using the number of days between visits to enable the use of time dummies. A program effect that reduces the days between visits therefore implies a higher frequency of visits. Figure 1 shows a histogram of days between visits for customers that return within a year. Customers are most likely to come back between 3 to 10 weeks after a visit. Because the data are right-censored, we run our analysis on the number of days between visits conditional on customers returning within 182 days (which is 26 weeks, or approximately half a year). Note we use this limit only for the descriptive analysis, and it is not necessary (and not used) for the HMM analysis.

*Figure 1. Histogram of Number of Days between Visits*
We estimate the model on frequency using the same two empirical specifications, equations (1) and (2), except that $Y_{ik}$ now represents the number of days between the $k^{th}$ and $(k + 1)^{th}$ visits for member $i$. We report the results in columns 1 and 2 of Table 4. The results are very similar between the two specifications with individual fixed effects (column 1) and segments based on the number of pre-program visits (column 2). The program reduces the days between visits by approximately 2–2.3 days, which reflects an approximately 3%–4% increase in the frequency of visits.

<table>
<thead>
<tr>
<th></th>
<th>Days between Visits</th>
<th>Attrition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td><strong>Program Start: Automatic Members</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program Start: Automatic Members</td>
<td>-2.3306$^*$</td>
<td>-1.9512$^{**}$</td>
</tr>
<tr>
<td></td>
<td>(1.4021)</td>
<td>(0.8521)</td>
</tr>
<tr>
<td>Automatic Members</td>
<td>-2.7616$^{***}$</td>
<td>-0.1548$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.5854)</td>
<td>(0.0549)</td>
</tr>
<tr>
<td>Number of Visits: 2</td>
<td>15.9652$^{***}$</td>
<td>-0.1575</td>
</tr>
<tr>
<td></td>
<td>(1.4656)</td>
<td>(0.1670)</td>
</tr>
<tr>
<td>Number of Visits: 3</td>
<td>10.3776$^{***}$</td>
<td>-1.2182$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(2.1886)</td>
<td>(0.1714)</td>
</tr>
<tr>
<td>Number of Visits: 4</td>
<td>-0.6420</td>
<td>-2.0313$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(2.3756)</td>
<td>(0.1764)</td>
</tr>
<tr>
<td>Number of Visits: 5</td>
<td>-10.6031$^{***}$</td>
<td>-2.5018$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(2.3175)</td>
<td>(0.1833)</td>
</tr>
<tr>
<td>Number of Visits: 6</td>
<td>-17.1614$^{***}$</td>
<td>-2.9297$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(2.1031)</td>
<td>(0.1965)</td>
</tr>
<tr>
<td>Individual Dummies</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Year+Week Dummies</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>15,734</td>
<td>15,734</td>
</tr>
</tbody>
</table>

Note: Cluster-robust standard errors $^{*}p<0.1; ^{**}p<0.05; ^{***}p<0.01$
The final component we consider is how the loyalty program influences customer attrition. Unlike the analysis with spending or visit frequency, we cannot use individual fixed effects for the attrition analysis, because attrition is not a repeated event for most customers. We therefore use the matched segments to control for customer heterogeneity. Given the non-contractual nature of our setting, attrition is not observed by the firm. For the purposes of the descriptive analysis, we select a time duration of inactivity, 182 days (half a year), after which the customer is considered to have attrited. We do not impose this assumption in our main HMM-based analysis in sections 3 and 4.

We run a logistic-regression model to estimate the program’s effect on the probability of attrition. The logistic-regression analysis follows the same form as in equation (2), except that \( Y_{ik} \) equals 1 if customer \( i \) does not revisit within 182 days after visit \( k \), and 0 otherwise. The results are shown in column 3 of Table 4. After controlling for the time trend by the year-week fixed effects and customer heterogeneity by the number of pre-program visits, we see that automatic members do indeed have a lower probability of attrition after the program introduction. Turning the parameter estimates into attrition rates using the logistic function, we take the sample average of all the other variables in the regression multiplied by their corresponding parameter estimates to get a baseline coefficient of -1.540. The baseline attrition is then \[
\frac{\exp(-1.540)}{1+\exp(-1.540)} = 17.7\%.
\] For automatic members after the program introduction, the attrition rate is estimated to be \[
\frac{\exp(-1.540-0.168)}{1+\exp(-1.540-0.168)} = 15.3\%.
\] Thus, the attrition rate reduces from 17.7% to 15.3% for automatic members after program introduction, which corresponds to a 13% reduction in relative terms.

To summarize, the loyalty program has only a negligible effect on spending per visit and leads to about a 3%–4% increase in the frequency of visits. This finding is consistent with the small effects found in prior literature (e.g., Sharp and Sharp, 1997; Lewis, 2004; Leenheer et al., 2007). However, the loyalty program has a large impact on reducing customer attrition, which has been missed in the prior literature on non-tiered loyalty programs. Because we cannot directly observe attrition in our non-contractual setting, we use a reasonable but somewhat arbitrary cutoff for when customer attrition would occur. In section 3, we propose a variant of an HMM that accounts for frequency and attrition in one unified framework, and
removes the need to create an arbitrary cutoff duration for attrition. We do not focus on changes in spending in our HMM setting given the negligible effects shown here.

3. Model Development

To estimate the effect of the loyalty program on overall customer value (and customer behaviors that drive value), we require a model that can jointly capture attrition and visit frequency, for which three key attributes are needed. First, to account for attrition, the model needs to distinguish between periods of time when a customer is completely dormant (i.e., has a zero probability of visiting the salon) and when a customer is still active (i.e., has a non-zero probability of visiting). Second, the probability of visiting the hair salon can depend on the time since the customer’s last visit, because hair services are typically consumed at some periodicity between visits (which can be heterogeneous across consumers). Third, the model needs to allow for customers potentially switching back and forth between being active and being dormant.

These three conditions can be accounted for in an HMM framework (e.g., Netzer et al., 2008; Fader et al., 2010; Schweidel et al., 2011; Schwartz et al., 2014), which is popular in the customer-lifetime-value literature. In our setting, the dependent variable, \( Y_{it} \), is an indicator variable that equals 1 if customer \( i \) visited the salon at time \( t \), and 0 otherwise. Hence, each customer’s outcomes can be represented as a vector of binary choices. An HMM posits that these outcomes are generated by the customer’s traversal through a sequence of discrete hidden (i.e., unobserved) states over time.

An HMM is defined by three primitives: (1) the initial probability of being in a given state \( k \) at time 0, which we denote as \( p(s_0 = k) \); (2) the state transition probability matrix that captures the chance of switching states (or remaining in the same state) at each time period; and (3) the state-dependent visit probability \( p(Y_{it} = 1|s_{it} = k) \). The joint likelihood of the data and state sequence \( f(Y_i, s_i) \) is defined in terms of these three primitives, where \( Y_i \) is a vector of all visit choices made over time by customer \( i \) and \( s_i \) is the corresponding vector of hidden states. The model parameters can be estimated by maximizing the
marginal likelihood function (which is obtained by summing $f(Y_i, s_i)$ over all feasible $s_i$). In a typical HMM, the marginal likelihood is efficiently computed using a forward algorithm that leverages two assumptions made by an HMM: (1) the first-order Markovian property that the probability of being in a given state depends only on the previous state, and (2) $Y_{it}$ is conditionally independent of $\{Y_{i1}, \ldots, Y_{i,t-1}\}$ and $\{s_{i1}, \ldots, s_{i,t-1}\}$ after conditioning on $s_{it}$. In other words, assumption (2) states that $s_{it}$ and $p(Y_{it} = 1|s_{it} = k)$ contain all the necessary information for the probability distribution of $Y_{it}$.

These assumptions will be violated by our model’s requirement to allow $Y_{it}$ to be a function of the time since the last visit, which we denote as $d_{it}$. In other words, $p(Y_{it} = 1|d_{it}, s_{it})$ can be a function of $d_{it}$ and not just $s_{it}$. At first glance, this issue seems to be easily resolved by allowing $p(Y_{it} = 1|d_{it}, s_{it})$ to be a non-parametric function of $d_{it}$ using a series of dummy variables for $d_{it} \in \{1,2,3,\ldots\}$. Such an approach is akin to a discrete hazard model if we replace the HMM with a single-state model. However, allowing for duration dependence via a discrete hazard function and allowing for state transitions to and from a dormant state results in identification challenges under a standard HMM framework. We therefore develop an extension of an HMM that resolves these challenges. We term our proposed model a duration-dependent HMM (DD-HMM). We present the proposed modeling framework in section 3.1. In section 3.2, we discuss how we use the model to estimate the treatment effects of the loyalty program.

3.1 DD-HMM framework

We propose a model with two states: active (A) and dormant (D). In the active state, the customer has some probability of visiting a salon, $p(Y_{it} = 1|d_{it}, s_{it} = A) > 0$. In the dormant state, the customer will never visit the salon, that is, $p(Y_{it} = 1|s_{it} = D) = 0$. For notational convenience, we define the discrete hazard of visiting conditional on being in the active state as $h_{it}(d_{it}) = p(Y_{it} = 1|d_{it}, s_{it} = A)$.

The customer has some probability of transitioning between states A and D. The probability of transitioning from state A to D is given by $\theta_{AD}(d_{it})$, which can be a function of duration since last visit. The probability of transitioning from state D to A is $\theta_{DA}$, which is not a function of duration, because no
visit occurs while in state D. Therefore, \( d_{it} \) can be interpreted as the time spent in the active state since the last visit. Customers start in state A at their first visit (because their initial visit must occur in the active state) such that \( p(s_{i0} = A) = 1 \).

The framework is illustrated graphically in Figure 2.

**Figure 2: Graphical Representation of the Duration-Dependent HMM (DD-HMM)**

Before introducing the joint likelihood \( f(Y_i, s_i) \), we discuss some necessary exclusion restrictions to identify the proposed DD-HMM. First, the transition from the active to the dormant state can only happen right after a visit. This is equivalent to setting \( \theta_{AD}(d_{it}) = \begin{cases} \theta_{AD} & \text{if } d_{it} = 1 \\ 0 & \text{if } d_{it} \geq 2 \end{cases} \). In other words, if a customer remains in the active state right after a visit \( (d_{it} = 1) \), he stays there until the next visit, at which point he gets another opportunity to transition to the dormant state (similar to Fader et al. 2005). In the dormant state, customers can return to the active state with probability \( \theta_{DA} \) in each time period. This restriction to only allow movement to the dormant state at \( d_{it} = 1 \) is needed because if transitioning from active to dormant could happen at any time (i.e., if \( \theta_{AD}(d_{it}) \) is an unrestricted function of \( d_{it} \)), the model could allow for multiple back-and-forth state-switching between two visits, which leads to identification issues.

Second, we impose the restriction of a weakly increasing hazard. In other words, the probability of visiting increases with the duration since the last visit for customers in the active state. This restriction is
reasonable in our empirical context because demand for hair services is driven by hair growth, which increases with time. Without this restriction, the model can rationalize a long period of inactivity with an eventually decreasing hazard function as well as being in the dormant state. The weakly increasing hazard allows us to separately identify the hazard function while in the active state from the state transition probabilities.

With these restrictions, the model considers a set of state sequences that could occur between two visits—and the dimensionality of this state sequence increases with the duration between visits. As an illustration, consider a customer who makes two visits spaced four time periods apart. The possible state sequences for this behavior are shown in Table 5. For each of those sequences, $d_{it}$ takes on a different value because only time spent in the active state counts toward the duration since the last visit. The likelihood function therefore needs to consider all possible state sequences that affect $d_{it}$.

<table>
<thead>
<tr>
<th>State Sequence</th>
<th>Duration $d_{it}$</th>
<th>Intuition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAA</td>
<td>4 periods</td>
<td>Always active since last visit</td>
</tr>
<tr>
<td>DAAA</td>
<td>3 periods</td>
<td>Dormant for one period and returned to active</td>
</tr>
<tr>
<td>DDAA</td>
<td>2 periods</td>
<td>Dormant for two periods and returned to active</td>
</tr>
<tr>
<td>DDDA</td>
<td>1 period</td>
<td>Dormant for three periods and returned to active</td>
</tr>
</tbody>
</table>

We now proceed to write down the likelihood function of the DD-HMM. As with an HMM, the DD-HMM has three components: the initial state distribution $p(s_{i0} = A)$, the state transition matrix defined by $\theta_{AD}$ and $\theta_{DA}$, and the state-dependent choice probability $P(Y_{it} = 1|d_{it}) = h_{it}(d_{it})$, where $Y_{it}$ equals 1 if customer $i$ chooses to visit at time $t$.

The initial state distribution $p(s_{i0} = A) = 1$ because we assume the customer enters the panel data with their first transaction when they are in the active state $A$. The initial visit does not feature in the likelihood function.
We operationalize the hazard function \( h_{it}(d_{it}) \) with a vector \( \beta \) of length \( K \). The first \( K - 1 \) parameters non-parametrically define the hazard rate for the first \( K - 1 \) weeks. The \( K \)th parameter represents the hazard rate for weeks \( K \) and beyond.\(^7\) To satisfy the increasing hazard restriction, we require that \( \beta_k < \beta_{k+1} \forall k \). For our empirical setting we chose \( K = 7 \) as higher dimensionality did not significantly improve model fit.

\[
h_{it}(d_{it}) = \Phi \left[ \sum_{k=1}^{K-1} \beta_k \cdot I(d_{it} = k) + \beta_K I(d_{it} \geq K) \right]. \tag{3}
\]

We define state sequence \( s^t_i \equiv \{ s_{i1}, \ldots, s_{it} \} \), a vector of states for customer \( i \) from time 1 to \( t \). The joint likelihood of the data \( \{ Y_{it} \} \) and \( s^T_i \) for an individual customer \( i \) is

\[
L(\theta_{DA}, \theta_{DA}, \beta | \{ Y_{it} \}, s^T_i) = \prod_{t=1}^{T} h_{it}(d_{it}(s^t_i))^{I(Y_{it}=1,S_{it}=A)} \cdot [1 - h_{it}(d_{it}(s^t_i))]^{I(Y_{it}=0,S_{it}=A)} \cdot (\theta_{DA})^{I(s_{lt-1}=A,S_{lt-1}=D,Y_{lt-1}=1)} \cdot (\theta_{DA})^{I(s_{lt-1}=D,S_{lt-1}=A)} \cdot (1 - \theta_{DA})^{I(s_{lt-1}=D,Y_{lt-1}=1)} \cdot (1 - \theta_{DA})^{I(s_{lt-1}=A,S_{lt-1}=A,Y_{lt-1}=1)}. \tag{4}
\]

The marginal likelihood is obtained by summing over the space of possible state sequences \( s^T_i \), and the likelihood over all customers is formed by taking the product of the individual customer likelihoods:

\[
L(\theta_{DA}, \theta_{DA}, \beta | \{ Y_{it} \}) = \sum_{q \in s^T_i} L(\theta_{DA}, \theta_{DA}, \beta | \{ Y_{it} \}, q). \tag{5}
\]

The estimation of the DD-HMM model requires changes to the forward algorithm used in HMM estimation, because the entire history of state transitions and outcomes between consecutive visits will matter in summing over the full set of paths that lead from the active state in which one visit occurred and the active state in which the next visit occurred. We describe a computationally efficient approach to estimating the DD-HMM in Web Appendix A, along with a simulation analysis showing parameter recovery.

---

\(^7\) The hazard function is assumed to be flat after \( K \) weeks, which helps constrain the dimensionality of the state space (akin to the hidden semi-Markov modeling literature, e.g., McShane et al. 2013).
3.2 Using DD-HMM to estimate program treatment effects

The DD-HMM provides an integrated approach to model customer attrition and visit frequency. We now describe how we use the model to estimate the effect of program membership on attrition, visit frequency, and overall customer value. To estimate how attrition probability (transition from active to dormant, \( \theta_{AD} \)) and visit frequency (captured by hazard parameters, \( \beta \)) change after program introduction, we allow parameters \( \theta_{AD} \) and \( \beta \) to differ before and after the program introduction. Mathematically, we define \( L_{P_{it}} \) as an indicator of whether the program has launched at time \( t \). As described in Web Appendix A, we set

\[
\theta_{AD}(L_{P_{it}}) = \theta_{AD,\text{before}} \cdot \mathbb{I}(L_{P_{it}} = 0) + \theta_{AD,\text{after}} \cdot \mathbb{I}(L_{P_{it}} = 1), \quad \text{and} \quad \beta(L_{P_{it}}) = \beta_{\text{before}} \cdot \mathbb{I}(L_{P_{it}} = 0) + \beta_{\text{after}} \cdot \mathbb{I}(L_{P_{it}} = 1).
\]

We are then able to obtain a difference-in-differences (DID) estimate of the program’s effect for each attrition and hazard parameter, which represents how much of the automatic members’ attrition and hazard parameters can be attributed to the program effect. In this approach, non-members serve as an important control for time trends that are unrelated to membership. To account for customer heterogeneity, we further segment automatic and non-members by the number of pre-program visits the customer made (six segments corresponding to one to six visits made before the loyalty program’s introduction). We then estimate separate models for each segment of customers, as shown in Table 6. Note that we run separate models for each segment, as well as separate models based on whether the customers are automatic members or non-members. Thus, our DID estimates are:

\[
\begin{align*}
\text{DID}(\theta_{AD,s}) &= (\theta_{AD,\text{after,auto},s} - \theta_{AD,\text{before,auto},s}) - (\theta_{AD,\text{after,non},s} - \theta_{AD,\text{before,non},s}) \\
\text{DID}(\beta_{k,s}) &= (\beta_{k,\text{after,auto},s} - \beta_{k,\text{before,auto},s}) - (\beta_{k,\text{after,non},s} - \beta_{k,\text{before,non},s})
\end{align*}
\]

---

8 We note the treatment is a regime change (launch of loyalty program) rather than a temporary benefit that expires.
9 Examples can include opening of competing stores or changes in hair style preferences (e.g., keeping one’s hair longer or shorter).
10 The number of customers with seven or more visits before the program introduction was too low to reliably estimate treatment effects. Using the customers with six or fewer visits account for over 60% of the cohort’s transactions.
Table 6: Estimated DD-HMM Models

<table>
<thead>
<tr>
<th>Segment (Customer Type)</th>
<th>Automatic Members</th>
<th>Non-members</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pre-program visit</td>
<td>DDHMM-1A</td>
<td>DDHMM-1N</td>
</tr>
<tr>
<td>2 pre-program visits</td>
<td>DDHMM-2A</td>
<td>DDHMM-2N</td>
</tr>
<tr>
<td>3 pre-program visits</td>
<td>DDHMM-3A</td>
<td>DDHMM-3N</td>
</tr>
<tr>
<td>4 pre-program visits</td>
<td>DDHMM-4A</td>
<td>DDHMM-4N</td>
</tr>
<tr>
<td>5 pre-program visits</td>
<td>DDHMM-5A</td>
<td>DDHMM-5N</td>
</tr>
<tr>
<td>6 pre-program visits</td>
<td>DDHMM-6A</td>
<td>DDHMM-6N</td>
</tr>
</tbody>
</table>

This approach gives us a flexible way to estimate the consumer heterogeneity. In other words, instead of allowing for a random effect on the HMM parameters to capture heterogeneity based on a distributional form, we allow the parameters to be completely flexible for each segment.

Based on the estimated parameters and DID effects, we are able to simulate discounted overall customer value over a five-year horizon for automatic members after program introduction in each segment. Discounting is applied at the weekly level using a nominal annual rate of 10% for the simulated individual-level customer visits that occur over five years. The discounted expected number of visits is then multiplied by $21, the typical price of a haircut when we calculate overall customer value.11

To simulate what these automatic members would have done had the program impact been zero, we subtract the DID estimates of the attrition and hazard parameters from the parameters for automatic members after program introduction. This then represents the appropriate counterfactual of having adjusted parameters that maintain the same difference between automatic- and non- members in the before-program period to also hold in the after-program period. From this automatic member “control” condition, we simulate the discounted overall customer value, and compare the change and lift in value by segment.

Further, we are also able to separate the relative effect of attrition and frequency on customer value by simulating the customer value under two further conditions – one in which only attrition effects are present (i.e., frequency effects are “turned off”) and one in which only frequency effects are present (i.e.,

---

11 We show in section 2 that spending per visit is relatively stable over time and customers
attrition effects are “turned off”). The lift due to each of these conditions provides an indication of whether attrition or frequency effects contribute more towards changes in customer value.

4. Model Results

In this section, we present the results from the estimated model. In section 4.1, we show that the DD-HMM parameters for automatic and non-members are generally similar before the program introduction. In section 4.2, we present DID estimates for attrition and visit frequency. In section 4.3, we present DID estimates for discounted customer value over a five-year time horizon. Other detailed DD-HMM results are found in Web Appendix B.

4.1 Pre-program DD-HMM parameters for automatic and non-members

Customers are segmented by pre-program visit frequency for both automatic and non-members. We show that the DD-HMM parameters in the pre-launch phase are similar between matched segments of automatic and non-members. While we do not need the members and non-members to act the same before the loyalty program is enacted because we are using a DID approach, which merely requires a common set of time effects, the finding that our parsimonious matching approach appears to account for most of the difference between automatic and non-members before the program introduction is reassuring. The key variables from the pre-program period include $\theta_{\text{AD,before}}$ (the transition probability from active to dormant) and $h_{\text{before,mem}}(d)$ (the hazard rate of visiting in week $d$ (or for $d=7$, week 7 or above) conditional on being in the active state and not having visited in the prior weeks, for automatic members ($\text{mem} = \text{auto}$) or non-members ($\text{mem} = \text{non}$)).

Table 7 presents the differences in these probabilities between automatic and non-members. We observe that only five of the 48 parameters are statistically significant (highlighted in bold in Table 7). None of the attrition probability parameters is statistically different between matched automatic and non-member segments, which suggests that non-members do not have a higher attrition probability than automatic
members before program introduction. Most of the hazard-rate parameters are not statistically different. Among the few statistically significant results, the effect is largest for week 5, where we observe that non-members in segments 3, 5, and 6 are 22%–35% less likely to make a visit exactly five weeks after their previous visit. Overall, the small differences suggest that our estimation strategy controls for most of the pre-program differences between automatic and non-members, and that matched non-members can be used as a reasonable control group.

### Table 7: Difference in DD-HMM Attrition and Hazard before Program Introduction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment (Pre-program visit frequency)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>( \theta_{AD, before, auto} ) - ( \theta_{AD, before, non} )</td>
<td>-0.012</td>
<td>0.037</td>
<td>0.019</td>
<td>-0.002</td>
<td>-0.011</td>
<td>-0.006</td>
</tr>
<tr>
<td>( h_{before, auto} (1) ) - ( h_{before, non} (1) )</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td>( h_{before, auto} (2) ) - ( h_{before, non} (2) )</td>
<td>0.000</td>
<td>-0.004</td>
<td>-0.002</td>
<td>0.000</td>
<td>-0.005</td>
<td>-0.009</td>
</tr>
<tr>
<td>( h_{before, auto} (3) ) - ( h_{before, non} (3) )</td>
<td>0.000</td>
<td>0.007</td>
<td>0.010</td>
<td><strong>0.023</strong></td>
<td>0.018</td>
<td>-0.005</td>
</tr>
<tr>
<td>( h_{before, auto} (4) ) - ( h_{before, non} (4) )</td>
<td>0.000</td>
<td>0.012</td>
<td><strong>0.026</strong></td>
<td>0.005</td>
<td>0.015</td>
<td>-0.019</td>
</tr>
<tr>
<td>( h_{before, auto} (5) ) - ( h_{before, non} (5) )</td>
<td>0.000</td>
<td>0.007</td>
<td><strong>0.029</strong></td>
<td>-0.001</td>
<td><strong>0.046</strong></td>
<td><strong>0.064</strong></td>
</tr>
<tr>
<td>( h_{before, auto} (6) ) - ( h_{before, non} (6) )</td>
<td>0.001</td>
<td>0.004</td>
<td>0.015</td>
<td>0.020</td>
<td>0.013</td>
<td>0.020</td>
</tr>
<tr>
<td>( h_{before, auto} (7) ) - ( h_{before, non} (7) )</td>
<td>0.001</td>
<td>0.004</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.030</td>
<td>0.017</td>
</tr>
</tbody>
</table>

*Note: Differences that are statistically significant (p-value < 0.05) are bolded.*

#### 4.2 Program’s effect on attrition, frequency, and overall customer value

In this section, we analyze the program’s effects on customer value, by first examining the program’s effect directly on the state transition variables and the visit hazard rates. We then calculate the overall effect of
the program on customer value, and break that effect down into how much comes from attrition and
frequency, demonstrating that most of the effect comes from attrition.

We begin by examining how the DD-HMM computes the posterior probability that a customer has
remained in the active state in the duration between visits, in order to build intuition for how the model
teases apart dormancy from infrequent visits while still active. We obtain these results using the posterior
probability of *not* having entered the dormant state as a function of inter-visit duration. Figure 3 shows this
posterior probability for the automatic members in the six different customer segments after the program
introduction. The results are similar for non-members.

Figure 3: Posterior Probability of Remaining in the Active State (without transitioning to dormant) as a
Function of Inter-visit Duration (in weeks) for the Six Customer Segments

Note: The curves going from left to right are in descending order of pre-program visit frequency. Note that
the curves for segments 5 and 6 are virtually the same.
Note that the posterior probability curve for each segment is based on the segment-specific estimated model (i.e., DDHMM-1A, …, DDHMM-6A). As a result, the differences in the posterior probability of (always) remaining active are driven by the variation in the estimated state transition matrix and hazard functions. Further, even though we show the posterior probability for always remaining active, the DD-HMM takes into account all other possible state sequences (which include the possibility that the customer became dormant and then became active again) in constructing the likelihood function.

We observe that as the pre-program visit frequency increases, the duration at which the model implies the customer likely becomes dormant decreases. The intuition is as follows: We can be more confident that a frequent customer who does not show up for some time is likely to have attrited, but a less frequent customer who has been absent for the same duration might still be active. This intuition is in line with previous work on non-contractual CLV modeling (e.g., Fader et al., 2010). In this sense, Figure 3 provides reassurance that the DD-HMM adapts the threshold for when a customer is considered to have attrited based on past behaviors.

We next discuss the direct DID estimates of the program effect on the attrition and visit frequency frequency. First, we examine how the program changed the attrition rate, which is measured as how much $\theta_{AD}$ is reduced for automatic members after program introduction. The estimated reduction in the attrition probability is given in Table 8. Overall, we find that the attrition rate decreases by 5.0 percentage points, by taking a weighted average of the segment-specific effect sizes. This amount corresponds to a 21.9% relative reduction in the attrition probability. This result is of a similar order of magnitude to that in the descriptive analysis of section 2, which found that the attrition rate dropped by 2.4 percentage points (from 17.7% to 15.3%), a 13% relative reduction. This overall attrition effect masks a considerable amount of heterogeneity. The program has the largest effect of reducing attrition on the least (one visit prior to program) and most (six visits prior to program) frequent customers. Additionally, the program appears to have some

---

12 This is calculated by taking the weighted average of the attrition effects (weighted by the proportion of automatic members in each segment as shown in Table 2) and dividing by the weighted average of the attrition probabilities in the absence of any effects
effect in reducing attrition for relatively infrequent customers (with two or three pre-program visits), though the effects are marginal in terms of statistical significance. We do not see an effect for moderately frequent customers (with four or five pre-program visits).

Table 8: DID Program Effect Estimate on Attrition (Probability of Transition from Active to Dormant State)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Program Estimate</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pre-program visit</td>
<td>-0.071</td>
<td>0.022</td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>2 pre-program visits</td>
<td>-0.047</td>
<td>0.029</td>
<td>0.104</td>
</tr>
<tr>
<td>3 pre-program visits</td>
<td>-0.046</td>
<td>0.024</td>
<td>0.051</td>
</tr>
<tr>
<td>4 pre-program visits</td>
<td>0.003</td>
<td>0.019</td>
<td>0.863</td>
</tr>
<tr>
<td>5 pre-program visits</td>
<td>-0.014</td>
<td>0.018</td>
<td>0.433</td>
</tr>
<tr>
<td>6 pre-program visits</td>
<td>-0.053</td>
<td>0.017</td>
<td><strong>0.002</strong></td>
</tr>
</tbody>
</table>

*Note: P-values less than 0.05 are in bold.*

These results suggest the following. First, segment 1 is composed of customers who have made only one pre-program visit to the salon and have experienced a hiatus of at least six months. Of those customers who do return after program launch, we find that automatic members have an attrition rate that is 7.1 percentage points lower than the rate for non-members. This finding suggests that the loyalty program reduces the chance of entering a long hiatus even for the segment 1 customers who will take a long time to receive a reward coupon at their rate of visits. Second, segment 6, which is composed of highly frequent customers, also experiences an absolute reduction in attrition of 5.3 percentage points. This finding suggests that the program is also effective at reducing periods of hiatus for these frequent customers, perhaps because these customers anticipate that they are likely to earn rewards. We can only conjecture on the absence of an effect for moderate-frequency customers. One reason may be that some of these customers are using a mix of hair-service options other than our focal salon chain. Because demand for hair services is roughly constant (because hair growth is linear in time), moderate-frequency customers at our focal salon may be
perusing a variety of competitive options and therefore are unlikely to experience behavior change in terms of attrition.

The loyalty program can also affect the frequency with which customers patronize the hair salon conditional on being in the active state, as measured by the hazard rate. These results are shown in Table 9. A negative value indicates that the hazard of visiting with the given duration has increased. We observe that segment 5 has a statistically significant and positive lift in the hazard for three out of seven hazard coefficients. Segments 2 and 6 have two statistically significant and positive coefficients. Our model-based analysis suggests that the frequency effect is also heterogeneous across segments – with larger effect sizes on high-frequency customers (segments 5 and 6).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment (Pre-program visit frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>( h(1) )</td>
<td>0.003</td>
</tr>
<tr>
<td>( h(2) )</td>
<td>\textbf{-0.020}</td>
</tr>
<tr>
<td>( h(3) )</td>
<td>\textbf{0.029}</td>
</tr>
<tr>
<td>( h(4) )</td>
<td>-0.010</td>
</tr>
<tr>
<td>( h(5) )</td>
<td>0.000</td>
</tr>
<tr>
<td>( h(6) )</td>
<td>0.019</td>
</tr>
<tr>
<td>( h(7) )</td>
<td>0.009</td>
</tr>
</tbody>
</table>

*Note: Differences that are statistically significant (p-value < 0.05) are bolded.*

Now that we have laid out the changes in the underlying key elements, we next synthesize how these elements add up in terms of increasing the total value the firm gets from its customers. We compute the effect of the program on the overall customer value over a five-year (or 260-week) horizon (similar to Fader et al., 2010), which we deem as a managerially relevant benchmark. We use the DD-HMM parameter estimates for each segment to forward-simulate customer visits on a weekly basis and compute the discounted expected number of visits from the simulated data. We apply discounting at the weekly level.
with a nominal annual discount rate of 10%.\textsuperscript{13} We then multiply the discounted expected number of visits by a $21 price per visit (which is the most common amount as the price of a basic hair service) to obtain the overall five-year discounted expected customer value.

We calculate the impact of the loyalty program using two sets of simulations for each segment. In the first set, we use the DD-HMM parameters for automatic members in the post-program introduction period, which includes the full effect of the program on attrition and frequency. In the second set, we use the same parameters as in the first set less the corresponding DID estimates for attrition probability and the hazard-function parameters, which represent a “control” condition of what the automatic members’ parameters would have been in the absence of the estimated program effects.

We show the results in Table 10. Column 1 reports the dollar value the firm would earn per customer in each segment if the firm did not have the loyalty program. Column 2 reports how much more the firm is able to extract from consumers in all of the segments, on average, because of the presence of the loyalty program. Finally, column 3 reports the percentage increase in the customer value as a result of the loyalty program. The program significantly improves value for most of the segments. The lower-frequency segments (1–3 pre-program visits) experience a lift in value between 22% and 25%. The highest-frequency segment (six pre-program visits) has the highest lift of 34.6%. The moderate-frequency segments have lower gains than the low- and high-frequency segments. The aggregate lift in customer value is 18.9%, which represents a substantial gain in average revenue per customer over five years.

These results show that program membership, while providing a healthy return for highly frequent customers (i.e., segment 6) over a five-year horizon, also provides strong returns for infrequent customers. The reason is that even if infrequent customers visit rarely to begin with, getting some occasional visits can add up relative to losing their business completely. Based on this analysis, conferring program membership

\textsuperscript{13} Discounting is done at the weekly level because this level is our unit of time in our model-based analysis. The weekly discount rate is therefore $10\%/52$ and is compounded on a weekly basis, to better reflect when cash flows are incurred.
appears to have an effect even for low-frequency customers who are less likely than high-frequency customers to earn a coupon for redemption.

Table 10: Program Effect and Lift Estimates on Discounted Expected 5-Year Customer Value

<table>
<thead>
<tr>
<th>Segment</th>
<th>Customer Value Without Program (Std. err.)</th>
<th>Change in Customer Value (Std. err.)</th>
<th>Lift in Customer Value (Std. err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$60.29 ($2.94)</td>
<td>$15.19*** ($3.03)</td>
<td>25.4%*** (5.8%)</td>
</tr>
<tr>
<td>2</td>
<td>$79.87 ($3.23)</td>
<td>$19.22*** ($4.08)</td>
<td>24.2%*** (5.6%)</td>
</tr>
<tr>
<td>3</td>
<td>$107.92 ($4.17)</td>
<td>$23.82*** ($4.21)</td>
<td>22.2%*** (4.4%)</td>
</tr>
<tr>
<td>4</td>
<td>$215.93 ($5.31)</td>
<td>$7.44 ($5.16)</td>
<td>3.5%*** (2.4%)</td>
</tr>
<tr>
<td>5</td>
<td>$297.36 ($6.18)</td>
<td>$34.52*** ($6.36)</td>
<td>11.6%*** (2.3%)</td>
</tr>
<tr>
<td>6</td>
<td>$300.88 ($6.89)</td>
<td>$104.00*** ($10.60)</td>
<td>34.6%*** (4.0%)</td>
</tr>
<tr>
<td>Overall</td>
<td>$118.51 ($1.63)</td>
<td>$22.35*** ($1.83)</td>
<td>18.9%*** (1.7%)</td>
</tr>
</tbody>
</table>

Note: *** indicates p-value < 0.05

To understand the relative impact of attrition versus frequency effects on overall customer value, we use a third set of parameters where we “turn off” the frequency effects (by setting them to their implied levels if the program had not existed) and leave only the attrition effects on the program to compute overall customer value. The results are shown in Table 11. They demonstrate a pattern similar to the one above, which is that the largest effects on value from attrition reduction occur for the most frequent customers, although we find large effects for infrequent customers, too. Note the total lift of the program from increased attrition is 13.8%, which represents most of the 18.9% total lift of the program.

We also conduct a similar analysis about the impact of the program on frequency effects. In this case, we use only the attrition parameters without the program and then compare how the average customer value changes when customers have the hazard rates of visits with the program compared to what would
happen if the program had not been implemented. The results are in Table 12. We observe that frequency-only effects have a null effect on customer value for most segments. Segment 5 has the highest lift due to frequency, and segment 4 has marginal statistical significance (p-value < 0.1). Segment 5 therefore has the highest proportion of its total lift (in Table 10) driven by frequency, as the attrition effect clearly dominates lift for other segments. The aggregate lift due to frequency-only effects is 3.7% (compared to the total lift of 18.9%).

<table>
<thead>
<tr>
<th>Segment</th>
<th>Customer Value without Program (Std. err.)</th>
<th>Change in Customer Value (Std. err.)</th>
<th>Lift in Customer Value (Std. err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$60.29 ($2.94)</td>
<td>$13.64*** ($2.99)</td>
<td>22.8%*** (5.6%)</td>
</tr>
<tr>
<td>2</td>
<td>$79.87 ($3.23)</td>
<td>$12.26*** ($3.83)</td>
<td>15.5%*** (5.1%)</td>
</tr>
<tr>
<td>3</td>
<td>$107.92 ($4.17)</td>
<td>$17.60*** ($3.89)</td>
<td>16.4%*** (3.9%)</td>
</tr>
<tr>
<td>4</td>
<td>$215.93 ($5.31)</td>
<td>-3.52 ($4.51)</td>
<td>-1.6% (2.1%)</td>
</tr>
<tr>
<td>5</td>
<td>$297.36 ($6.18)</td>
<td>$16.92*** ($6.02)</td>
<td>5.7%*** (2.1%)</td>
</tr>
<tr>
<td>6</td>
<td>$300.88 ($6.89)</td>
<td>$92.93*** ($11.00)</td>
<td>31.0%*** (4.2%)</td>
</tr>
<tr>
<td>Overall</td>
<td>$118.51 ($1.63)</td>
<td>$16.33*** ($1.75)</td>
<td>13.8%*** (1.6%)</td>
</tr>
</tbody>
</table>

Note: *** indicates p-value < 0.05

In summary, we see the program significantly increases the customer value for the firm, obtaining an 18.9% lift. Most of this effect comes from attrition, whereas a much smaller proportion comes from the increased frequency of visits. We also note that the total effect of the program on customer value, as measured in Table 10, is slightly larger than the effect of the program reported in Table 11 (attrition) and Table 12 (frequency). The reason is that a complementarity exists in attrition and frequency in increasing
customer value: Increased attrition becomes more valuable if customers come more frequently; Put another way, the value of increased frequency is only beneficial if the customers actually stay.

Table 12: Frequency-Only Estimates on Discounted Expected 5-Year Customer Value

<table>
<thead>
<tr>
<th>Segment</th>
<th>Customer Value without Program (Std. err.)</th>
<th>Change in Customer Value (Std. err.)</th>
<th>Lift in Customer Value (Std. err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$60.29 ($2.94)</td>
<td>$0.26 ($2.50)</td>
<td>0.5% (4.2%)</td>
</tr>
<tr>
<td>2</td>
<td>$79.87 ($3.23)</td>
<td>$5.47 ($3.17)</td>
<td>6.9% (4.2%)</td>
</tr>
<tr>
<td>3</td>
<td>$107.92 ($4.17)</td>
<td>$4.08 ($4.07)</td>
<td>3.9% (3.9%)</td>
</tr>
<tr>
<td>4</td>
<td>$215.93 ($5.31)</td>
<td>$9.95 ($5.33)</td>
<td>4.6% (2.5%)</td>
</tr>
<tr>
<td>5</td>
<td>$297.36 ($6.18)</td>
<td>$16.72*** ($7.15)</td>
<td>5.6%*** (2.5%)</td>
</tr>
<tr>
<td>6</td>
<td>$300.88 ($6.89)</td>
<td>$5.90 ($7.24)</td>
<td>2.0% (2.4%)</td>
</tr>
<tr>
<td>Overall</td>
<td>$118.51 ($1.63)</td>
<td>$4.43*** ($1.57)</td>
<td>3.7%*** (1.3%)</td>
</tr>
</tbody>
</table>

Note: *** indicates p-value < 0.05

Our measured impact of the loyalty program is much higher than the impact other papers studying these programs have found, and a significant reason for this is that other papers have not been able to study the impact of the program on attrition rates, which we find is the main driver of the loyalty programs’ value.

5. Discussion and Conclusion

In this study, we estimate the effect of a non-tiered loyalty program on customer value over a five-year time horizon, and decompose the drivers of this change in customer value into effects from attrition, visit frequency, and monetary spending. We show through careful descriptive analysis that the program has a null effect on amount spent per visit. Both the descriptive analysis and a model-based approach demonstrate the program has a modest effect on visit frequency and a large effect on attrition prevention. Although we
believe our model provides the best measures of the frequency and attrition, the finding that the magnitudes of the descriptive and modeled results are similar is reassuring because it suggests the effects come directly from the data, and are not imposed by the model structure. Ultimately, we find that the overall customer value over a five-year horizon improves by 18.9%, primarily driven by the higher attrition rates as a result of the program.

We also show that visit-frequency and attrition effects are highly heterogeneous. Visit-frequency effects are highest for the most frequent customer segments in our dataset. Attrition effects are strongest for low-frequency and high-frequency customer segments, but low for moderate-frequency customers. This finding, in turn, reflects the high customer-value lifts for customers on the low and high end of visit frequency but more modest lifts effects for those in the middle.

Our effect sizes show that a non-tiered loyalty program can indeed enhance revenue (and therefore profit) even though only a small minority of customers (i.e., the highest-frequency ones) actually earn and redeem the $5 off coupons. As a result, enrolling a customer as a member reduces attrition for the firm—at very little cost, because only 2% of customers redeem a rewards coupon. Further, even low-frequency customers show a behavior change in terms of attrition when enrolled in the program, suggesting the benefit of the program can extend beyond economic factors that relate only to rewards redemption.

Our study makes three significant contributions to the literature. First, we find a larger effect size for a non-tiered loyalty program, and we find this larger effect size is attained by including the potential effects of attrition—an aspect that extant literature has largely been silent on. Second, we leverage variation in our data that minimizes selection biases to the extent possible, which is aided significantly by having data from the customers both before and after the loyalty program is introduced. Third, separating attrition and visit-frequency effects is non-trivial because attrition is unobserved in the non-contractual setting, which is the case in our empirical context. Hence, we implement a modeling approach that is an extension of a Hidden Markov Model, but one that allows for an increasing hazard function while the customer actively considers visiting the hair salon, and a zero probability of a visit when the customer is in a dormant state. Our model therefore allows for a careful separation of these effects to better understand how much
the program affects attrition and frequency, and to use simulations based on estimated parameters to compute the five-year lift in customer value. Our modeling approach can be seen as a demonstration of how the customer-lifetime-value literature can be applied to a program-evaluation problem—in this case, of measuring the impact of a new loyalty program.
References


Web Appendix A: Estimation and Simulation Study of DD-HMM

Model Estimation

We estimate the DD-HMM described in Section 3 of the paper using Maximum Likelihood Estimation (MLE). A standard HMM could be estimated using the forward algorithm which integrates over the possible state sequences to arrive at an overall data likelihood for a given set of parameters. The forward algorithm works because of the first-order Markovian property of a standard HMM – that only the previous state affects the current state. However, as illustrated in Table 5, the DD-HMM also features duration since last visit \( d_{it} \) as a state variable – and this variable does not accumulate when in the dormant state. As a result, it becomes necessary to integrate over possible values of \( d_{it} \) in the likelihood function which requires keeping track of more than just the previous state (Active or Dormant).

We therefore develop a modification of the forward algorithm that exploits the fact that each customer’s data can be represented as a series of intervisit durations (after the initial visit) and a survival period at the end which represents how long the customer has not yet visited. We therefore need to take into account all possible state sequences that can occur for a given intervisit or survival duration. The number of such possible sequences increases rapidly as the intervisit duration increases.

For computational efficiency, we pre-compute the log likelihood of various components of the likelihood function for different intervisit and survival times so that these are not re-computed for each individual. What helps our model be computationally efficient is that parameters for a given customer segment are homogeneous – i.e., there are before- and after- program DD-HMM parameters that apply to all customers within a segment, while heterogeneity is flexibly captured by estimating separate models for each segment.

Model parameters for a given customer segment

The parameters for a given customer segment (within either automatic- or non- members) are the set of state transition and hazard parameters before- and after- program introduction.

\[
\theta_{AD}(LP_{it}) = \theta_{AD,before} \cdot I(LP_{it} = 0) + \theta_{AD,after} \cdot I(LP_{it} = 1), \quad \text{and} \quad \beta(LP_{it}) = \beta_{before} \cdot I(LP_{it} = 0) + \beta_{after} \cdot I(LP_{it} = 1).
\]

The transition probability going from dormant to active state \( \theta_{DA} \) is not a function of program introduction as there is not enough variation in the data to estimate these before- and after- the program.

For the hazard parameters, we have \( K \) non-parametric duration dummies and the \( K^{th} \) parameter captures a constant hazard for a duration that is \( K \) weeks or more since the last visit. \( K \) is therefore a modeling choice and we chose \( K = 7 \) for our empirical model as higher \( K \) did not substantially improve model fit.

Simulation study

To demonstrate the empirical recovery of model parameters, we designed a simulation study in which \( K = 7 \) and we chose model parameters from one of the actual estimated DD-HMM’s (segment 5 of automatic members) in Section 4. Below we show the simulated and recovered parameters for a simulation with 126 customers over 120 time periods (weeks), with a “regime change” after 30 time periods similar to our program introduction. We simulate 25 different data sets with segment 5’s automatic members’ parameters...
and compute the mean and standard deviation of the estimated parameters from the set of simulations as shown in the below table. As can be seen, parameter recovery is robust, and the true parameters lie in the 95% interval for all parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{DA,auto}$</td>
<td>0.015</td>
<td>0.016</td>
<td>0.002</td>
</tr>
<tr>
<td>$\theta_{AD, before, auto}$</td>
<td>0.047</td>
<td>0.047</td>
<td>0.010</td>
</tr>
<tr>
<td>$h_{before, auto}(1)$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$h_{before, auto}(2)$</td>
<td>0.010</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td>$h_{before, auto}(3)$</td>
<td>0.047</td>
<td>0.048</td>
<td>0.012</td>
</tr>
<tr>
<td>$h_{before, auto}(4)$</td>
<td>0.089</td>
<td>0.087</td>
<td>0.016</td>
</tr>
<tr>
<td>$h_{before, auto}(5)$</td>
<td>0.159</td>
<td>0.149</td>
<td>0.013</td>
</tr>
<tr>
<td>$h_{before, auto}(6)$</td>
<td>0.216</td>
<td>0.195</td>
<td>0.025</td>
</tr>
<tr>
<td>$h_{before, auto}(7)$</td>
<td>0.270</td>
<td>0.242</td>
<td>0.017</td>
</tr>
<tr>
<td>$\theta_{AD, after, auto}$</td>
<td>0.108</td>
<td>0.117</td>
<td>0.013</td>
</tr>
<tr>
<td>$h_{after, auto}(1)$</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>$h_{after, auto}(2)$</td>
<td>0.006</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>$h_{after, auto}(3)$</td>
<td>0.023</td>
<td>0.023</td>
<td>0.005</td>
</tr>
<tr>
<td>$h_{after, auto}(4)$</td>
<td>0.125</td>
<td>0.124</td>
<td>0.012</td>
</tr>
<tr>
<td>$h_{after, auto}(5)$</td>
<td>0.204</td>
<td>0.209</td>
<td>0.014</td>
</tr>
<tr>
<td>$h_{after, auto}(6)$</td>
<td>0.287</td>
<td>0.309</td>
<td>0.016</td>
</tr>
<tr>
<td>$h_{after, auto}(7)$</td>
<td>0.394</td>
<td>0.439</td>
<td>0.023</td>
</tr>
</tbody>
</table>
Web Appendix B: Full Set of DD-HMM parameter estimates

We use Maximum Likelihood Estimation (MLE) to obtain parameter estimates for each DD-HMM. We present the full set of parameter estimates for the models for automatic (Table B1) and non-members (Table B2) below. We present all parameters in terms of probabilities (taking the normal CDF transform from the Probit space) for more intuitive comparison.

Table B1: DD-HMM Parameters for Automatic Members

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment (Pre-program visit frequency)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{DA,auto}$</td>
<td>0.003</td>
<td>0.005</td>
<td>0.006</td>
<td>0.009</td>
<td>0.015</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>$\theta_{AD, before, auto}$</td>
<td>0.958</td>
<td>0.378</td>
<td>0.169</td>
<td>0.087</td>
<td>0.047</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{1, before, auto})$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{2, before, auto})$</td>
<td>0.000</td>
<td>0.010</td>
<td>0.004</td>
<td>0.004</td>
<td>0.010</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{3, before, auto})$</td>
<td>0.000</td>
<td>0.031</td>
<td>0.027</td>
<td>0.043</td>
<td>0.047</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{4, before, auto})$</td>
<td>0.000</td>
<td>0.051</td>
<td>0.062</td>
<td>0.061</td>
<td>0.089</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{5, before, auto})$</td>
<td>0.000</td>
<td>0.062</td>
<td>0.083</td>
<td>0.105</td>
<td>0.159</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{6, before, auto})$</td>
<td>0.001</td>
<td>0.063</td>
<td>0.106</td>
<td>0.145</td>
<td>0.216</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{7, before, auto})$</td>
<td>0.001</td>
<td>0.063</td>
<td>0.110</td>
<td>0.179</td>
<td>0.270</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td>$\theta_{AD, after, auto}$</td>
<td>0.231</td>
<td>0.172</td>
<td>0.151</td>
<td>0.111</td>
<td>0.108</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{1, after, auto})$</td>
<td>0.005</td>
<td>0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{2, after, auto})$</td>
<td>0.008</td>
<td>0.005</td>
<td>0.005</td>
<td>0.004</td>
<td>0.006</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{3, after, auto})$</td>
<td>0.068</td>
<td>0.036</td>
<td>0.024</td>
<td>0.032</td>
<td>0.023</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{4, after, auto})$</td>
<td>0.082</td>
<td>0.081</td>
<td>0.050</td>
<td>0.083</td>
<td>0.125</td>
<td>0.193</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{5, after, auto})$</td>
<td>0.114</td>
<td>0.137</td>
<td>0.104</td>
<td>0.091</td>
<td>0.204</td>
<td>0.304</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{6, after, auto})$</td>
<td>0.133</td>
<td>0.139</td>
<td>0.146</td>
<td>0.161</td>
<td>0.287</td>
<td>0.410</td>
<td></td>
</tr>
<tr>
<td>$\Phi(\beta_{7, after, auto})$</td>
<td>0.133</td>
<td>0.139</td>
<td>0.180</td>
<td>0.246</td>
<td>0.394</td>
<td>0.427</td>
<td></td>
</tr>
</tbody>
</table>
DID estimates are obtained parameter by parameter, by taking the difference between post and pre parameters.

These parameters are used to compute DID estimates by taking the difference between the after and before automatic member parameter, and the after and before non-member parameter. In order to compute what the automatic member’s parameter would have been in the absence of a program effect, we subtract the DID estimate from the automatic member’s after parameter. This then allows us to simulate a variety of conditions to obtain customer value:

- What an automatic member generates in customer value (based on the full effect of the program)
- What an automatic member generates if attrition and frequency effects are zero
- What an automatic member generates if attrition effects are zero but frequency effects remain
- What an automatic member generates if attrition effects remain but frequency effects are zero

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment (Pre-program visit frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\theta_{DA,\text{non}}$</td>
<td>0.002</td>
</tr>
<tr>
<td>$\theta_{AD,\text{before,non}}$</td>
<td>0.970</td>
</tr>
<tr>
<td>$\Phi(\beta_{1,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{2,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{3,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{4,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{5,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{6,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Phi(\beta_{7,\text{before,non}})$</td>
<td>0.000</td>
</tr>
<tr>
<td>$\theta_{AD,\text{after,non}}$</td>
<td>0.315</td>
</tr>
<tr>
<td>$\Phi(\beta_{1,\text{after,non}})$</td>
<td>0.002</td>
</tr>
<tr>
<td>$\Phi(\beta_{2,\text{after,non}})$</td>
<td>0.027</td>
</tr>
<tr>
<td>$\Phi(\beta_{3,\text{after,non}})$</td>
<td>0.039</td>
</tr>
<tr>
<td>$\Phi(\beta_{4,\text{after,non}})$</td>
<td>0.092</td>
</tr>
<tr>
<td>$\Phi(\beta_{5,\text{after,non}})$</td>
<td>0.114</td>
</tr>
<tr>
<td>$\Phi(\beta_{6,\text{after,non}})$</td>
<td>0.114</td>
</tr>
<tr>
<td>$\Phi(\beta_{7,\text{after,non}})$</td>
<td>0.123</td>
</tr>
</tbody>
</table>