The Impact of Subscription Programs on Customer Purchases

Raghuram Iyengar, Young-Hoon Park, and Qi Yu*

January 2020

* Raghuram Iyengar is Miers-Busch W'1885 Professor and Professor of Marketing at the Department of Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104; email: riyengar@wharton,upenn.edu. Young-Hoon Park is Sung-Whan Suh Professor of Management and Professor of Marketing at the Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853; email: yp34@cornell.edu. Qi Yu is a PhD student in Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104; email: qiyu@wharton.upenn.edu. All authors contributed equally and are listed in alphabetical order. We thank several managers at the participating company for their efforts in making this research possible. We thank Ryan Dew, Pinar Yildirim, Christophe Van den Bulte and seminar participants at 2017 and 2018 Marketing Science Conference, 2019 Choice Symposium, University of Chicago, University of Michigan at Ann Arbor, and University of Texas at Austin for their comments.

The Impact of Subscription Programs on Customer Purchases

Abstract

Subscription programs are increasingly popular among a wide variety of retailers. These types of programs give members access to a set of exclusive benefits for a fixed fee upfront. In this paper, we examine the causal effect of adopting a subscription program on subsequent customer behavior using a unique panel data from a company that launched a subscription program. To account for self-selection and identify the individual-level treatment effects, we combine a difference-in-differences approach with a generalized random forest that matches each member of the program with comparable non-members. We find the subscription leads to a large increase in customer purchases. The effect is economically significant, persistent over time, and heterogeneous across customers. Interestingly, only one-third of the effect on customer purchases is due to the economic benefits of the subscription program and the remaining is attributed to becoming a member per se. We provide evidence that members experience a sunk cost fallacy due to the upfront payment that subscription programs entail. We discuss the implications of our findings for customer retention and subscription services.

Keywords: Subscription business, Retailing, E-Commerce, Causal inference, Machine learning, Generalized random forest, Sunk cost fallacy

1. Introduction

The ability to retain and develop customers is critical to the success of a business. Toward this end, retailers are increasingly turning to subscription programs, which are designed to keep customers engaged by giving access to exclusive benefits for a fee upfront.[1] For example, Amazon Prime offers members unlimited free shipping, audio and video content, as well as member-exclusive discounts for an upfront payment of $119 per year. Many other retailers, such as Barnes & Noble, Sephora, and Alibaba have such programs as well, and benefits range from unlimited free shipping to member-only discounts and additional loyalty points.

As the relevance and popularity of subscription programs grows, it is of managerial interest to examine the causal effect of such programs on customer behavior and investigate underlying drivers for their success. For instance, an industry report speculates that Amazon Prime is quite successful, as members spend $1,300 per year, which is almost double the average non-member's annual spending of $700.[2] Yet, the reported difference in spending between members and non-members may arise due to several reasons. First, members likely self-select into the subscription. The naïve comparison in spending described above likely over-estimates the effect of the program, as customers who expect to make more purchases in the future are more likely to join the subscription. Second, members may actually change their purchase behavior due to the economic benefits they receive after joining the program. Third, mere membership can also bring value and change member behavior, for instance, by leading them to form a new consumption habit or to feel enhanced status. The industry report cited above

[1] We distinguish subscription programs from stand-alone subscription services (e.g., Stitch Fix, Birchbox) that provide subscribers new items or personalized experiences periodically. We focus on a setting where a subscription is initiated by an existing non-contractual business and provides members exclusive benefits beyond those available to regular customers (i.e., non-members). See Section 2 for a discussion on different types of subscriptions.
[2] "Here's How Much Amazon Prime Customers Spend Per Year," available at http://fortune.com/2017/10/18/amazon-prime-customer-spending/ (last accessed December 8, 2019).

indicates that 82% of Prime members shop with Amazon even though the price is lower elsewhere, suggesting that the impact of subscription programs can go beyond the offered economic benefits. From a theoretical perspective, it is important to separate the impact of the latter two components (i.e., economic benefits and membership per se) on customer behavior. While the economic effects of a program may be specific to its features, the underlying psychological drivers for why becoming a member of a program per se can change behavior are likely to be generalizable to other contexts. Determining the relative contribution of the two components is substantively important as well for improving the design of programs. As an extreme scenario, if additional sales are generated only by reducing (effective) prices, the program might negatively affect a firm's performance in the long term (e.g., Raghubir 2004).[3]

The purpose of this paper is to take a first step towards assessing the causal impact of customers joining a subscription program on their subsequent purchase behavior. We hereafter refer to the impact as the total treatment effect of the program. We also seek to separately identify the total effect of the program on purchase behavior that is due to becoming a member per se (which we hereafter refer to as the net treatment effect of the program) and that is due to the economic benefits offered in the program. Specifically, we are interested in addressing the following questions: Does the subscription program generate value for a firm? Is the subscription program effective in inducing customers to change their subsequent behavior because of the membership itself and/or the economic benefits associated with the program? How does the effect (both total and net) vary across customers and over time? What are the underlying drivers of any documented effects? We address these questions in close collaboration with a company

[3] For instance, Movie Pass, a subscription service that offered its members one free movie each day for the price of $9.95 per month, managed to attract more than 2 million subscribers but failed to build a deeper relationship with the customers. The company reported a loss of $266 million in 2018, and ended the subscription program in September 2019.

that launched a subscription program on its online website in 2015. The program offers members a few exclusive benefits for an upfront fee of $50 per year. Our data contains individual-level transactions before and after the launch of the program, and other information on various components in the program, thus presenting an appealing context to examine the causal effect of the program on purchase behavior.

A key concern while estimating the impact of a subscription program on purchase behavior arises from the lack of random assignment. We exploit the panel structure of our data and rich information on customer characteristics and rely on a quasi-experimental design to control for self-selection and identify the causal effect at the individual level. Specifically, our baseline model uses a difference-in-differences (DD) specification (Angrist and Pischke 2008) that controls for unobserved individual-fixed and time-fixed effects to estimate the total treatment effect on customer behavior. In addition, to enhance the comparability between members and non-members, we create a weighted set of neighboring observations for each customer based on a large set of observed characteristics following a generalized random forests procedure (Athey et al. 2019) and estimate the DD model using the weighted sample. The combination of the DD approach and the generalized random forests procedure alleviates several concerns with the former and is robust to selection bias based on observed as well as time-invariant unobserved characteristics. It also provides individual-level estimates of the treatment effect. Within this framework, we also quantify the net treatment effect of the program by controlling for the differential marketing mix each customer received. The individual-level estimates of the total and net treatment effects allow us to get a richer understanding of the impact of the program and its underlying drivers.

We find subscription is effective in lifting sales. On average, members increase their purchases by about $27 per month over a 12-month period after joining the program. As the average purchase amount per month is about $12 before subscription, the total effect of the program is economically significant and persistent over time. It keeps members more engaged in terms of frequency and variety in their purchases. Interestingly, only one-third of the total effect on purchase amount is due to the economic benefits of the program and the remaining two-thirds (about $18 out of $27) is attributed to becoming a member per se. There is also a large variation in the treatment effect across customers. Our main findings are robust to different outcomes of purchase behavior (e.g., including offline sales), different samples of the treated group, and potential confounding effects of self-selection and unobservables.

We investigate the potential drivers at work that explain our findings. We find that, in addition to psychological underpinnings (e.g., habit, status, affect) documented in the context of other types of membership programs (e.g., loyalty programs), a unique feature of subscription programs helps the sales lift: As customers pay a fee upfront in exchange for future benefits, they experience sunk cost fallacy in which they increase their purchases to justify their subscription decisions, even though the upfront fee is sunk (e.g., Thaler 1980, Arkes and Blumer 1985). We provide evidence supporting this mechanism.

Our paper is related to several streams of research. We contribute to the literature on subscription programs. Recently, McCarthy et al. (2017) develop a framework for valuing subscription-based firms, and Datta et al. (2017) study how the adoption of music streaming subscription affects listening behavior. Our paper makes both substantive and theoretical contributions to this nascent literature. Substantively, existing literature focuses on replenishment and curation subscriptions (e.g., McCarthy et al. 2017), which are in the form of standalone services. Our research extends this literature by studying a program initiated by an existing non-contractual business. Theoretically, we add to the literature that studies the underpinnings of subscription programs. We document a novel mechanism through which subscription programs can work (i.e., sunk cost fallacy). Because customers pay a fee upfront, they increase their purchases to take advantage of the program benefits in order to justify their subscription decisions. As the upfront fee is a feature common to subscription programs, we believe our results have broad implications that the effect of a subscription program can indeed go beyond the economic benefits it offers.

We also add to the literature on membership programs. Firms across a wide array of industries have long been using loyalty programs to reward repeat purchases, and there is extensive research on these types of programs. Some studies find loyalty programs can increase customer lifetime value and share of wallet (e.g., Lal and Bell 2003, Liu 2007, Kopalle et al. 2012). Others find no or weak evidence loyalty programs are effective (e.g., Hartmann and Viard 2008). Several researchers have documented loyalty programs can lead to the development of habitual consumption (e.g., Wood and Neal 2009), enhance members' perception of status (e.g., Drèze and Nunes 2009), and induce positive affect (e.g., Leenheer et al. 2007). This paper contributes to this literature by using quasi-experimental data to measure the causal effect of a subscription program on customer behavior.

Our findings suggest that customers behave in a boundedly rational manner, which adds to empirical evidence for such behavior found in the lab and in other field settings. The sunk cost fallacy has implications in a variety of contexts, and extensive evidence of it has been found in the lab (Thaler 1980). There are relatively few studies, however, providing evidence for the sunk cost fallacy in the field. For example, Arkes and Blumer (1985) conduct a field experiment and find the attendance rate is positively correlated with the price of theater tickets. Ho et al. (2017) find evidence for the sunk cost fallacy in the Singapore automobile markets where there is heterogeneity among consumers with regard to the payment for obtaining a government license to purchase a car. And the driving time is positively correlated with the price paid. We extend this literature by showing that subscription programs also induce the sunk cost fallacy and contribute to our understanding of consumer behavior in the field using observational data.

The remainder of the paper is organized as follows. Section 2 gives an overview of subscription business. Sections 3 and 4 discuss the data and methodology. Section 5 presents the results and discusses possible explanations for the effects. Section 6 presents several robustness checks. We conclude with directions for future search in Section 7.

2. Subscription Business

A subscription-based business is one in which a customer pays a fee to have access to products or services. Rather than selling products one at a time, a subscription offers periodic (e.g., monthly, yearly) use or access to products or services. Thus, a one-time purchase of a subscription can lead to recurring sales, and a predictable stream of revenues from subscribers. Pioneered by the likes of newspapers and magazines, more products and services are being offered through subscriptions than ever before. For instance, business-to-consumer subscription businesses attracted more than 11 million subscribers in 2017 in the U.S., and the industry as a whole has been growing at a staggering rate of 200% annually since 2011.[4]

[4] "Subscription Business Are Booming. Here's How to Value Them," available at https://hbr.org/2017/12/subscription-businesses-are-booming-heres-how-to-value-them (last accessed December 8, 2019).

Despite sharing a common feature of offering the use of or access to products or services for a fee, subscription-based businesses appear in many different formats. Existing business-to-consumer subscription business in retail can be broadly categorized into three types: replenishment, curation, and access.[5]

Replenishment subscriptions allow consumers to automate the purchase of commodity items, such as razors, diapers, and vitamins. With automatic replenishment, customers no longer need to take the time and effort to repeat the order themselves. Customers benefit from this type of subscription because it allows them to save time and money on each transaction. Examples include Dollar Shave Club, Gillette on Demand, and Rituals. Curation subscriptions seek to delight by providing new items or personalized experiences in such categories as apparel, beauty, and food. Examples include Stitch Fix, Birchbox, and Blue Apron.

In this paper, we focus on the third type of subscriptions, access subscriptions, which allow consumers to gain exclusive access or member-exclusive benefits. Access subscriptions have attracted substantial interest among more established retailers, as compared to the first two types of subscriptions, which are mostly launched by start-ups fueled by venture capital investments. Examples include Amazon (Prime), Barnes & Noble (B&N Membership), Sephora (Flash), Alibaba (88VIP), etc. These subscriptions differ from the other two described above in that a subscription is initiated by an existing non-contractual business and provides members exclusive benefits beyond those available to regular customers (i.e., non-members). Thus, these kinds of programs have a very wide appeal, as they can be adopted by nearly all business-to-

[5] There are more than 2,000 consumer-focused subscription businesses capitalizing on customers' diverse tastes in a wide range of categories. There are also hundreds of companies with more unorthodox products catering to the "long tail" of consumer tastes, including Harry Potter toys and survivalist products. As subscription-based businesses continue to grow, we recognize that the three categories described may not be sufficient to capture the diversity. See "Thinking Inside the Subscription Box: New Research on E-commerce Consumers," available at https://www.mckinsey.com/industries/high-tech/our-insights/thinking-inside-the-subscription-box-new-research-on-ecommerce-consumers (last accessed December 8, 2019).

consumer firms. While there are many variations of such programs in practice, benefits offered to members typically fall into two broad categories: unlimited use of a service (e.g., free shipping) and access to member-exclusive offers. Amazon (Prime) and Barnes & Noble (B&N Membership), for example, offer both types of benefits to their members while Sephora's Flash offers only unlimited free shipping. Firms also vary by the type of member-exclusive offers. Barnes & Noble provides members with exclusive offers only for purchasing products, while Amazon and Alibaba (88VIP) offer exclusive digital content to their members as well as member-only benefits related with product purchases. For the remainder of the paper, we refer to access subscriptions as subscriptions for brevity and use the terms subscription program and program interchangeably.

While there is anecdotal evidence suggesting the commercial success of subscriptions, no study has evaluated whether they indeed lead to incremental revenues from members as compared to non-members, who can purchase from a firm without being subscribers. It is also unclear whether the subscription program is effective in inducing customers to change their subsequent behavior because of the membership itself or the economic benefits associated with the program. For the former, mere membership can bring value to customers and influence their behavior. For instance, membership to an exclusive country club can bestow status and change purchase patterns.[6] For the latter, as discussed above, subscription programs usually come with some form of economic benefits. Our research aims to fill in this gap.

3. Empirical Context and Data

[6] "The In Crowd: Inside Boston's Elite Country Clubs," available at https://www.bostonmagazine.com/news/2018/09/11/in-crowd-country-clubs/ (last accessed on December 8, 2019).

We obtained the data for our empirical analysis from a retailer in Asia. The retailer sells a wide range of beauty products (e.g., skin care, make-up) and has both a brick-and-mortar and online presence. The retailer that we collaborated with launched a subscription program on its e-commerce website in December 2015. The launch of the program and its benefits were communicated to online customers through mass emails and on their website. The program provided both unlimited use of a service and access to member-exclusive offers. Specifically, for a subscription fee of $50 per year, members had access to unlimited free shipping with no minimum purchase requirements. Members also had access to several exclusive offers. Upon joining the program, customers were provided with a $50 gift card that could be used for purchases online with no restrictions. Members also received a $3 gift card per month for online purchases during the month. Free samples were offered to members monthly with a purchase online. Finally, several products were occasionally coupled with member-exclusive discounts. Similar to other subscription programs, these benefits were offered beyond the first year of subscription rollout.

Our data include all 10,811 customers who joined the program between December 2015 and February 2017. On average, 720 customers joined the program per month. The monthly number of members has a moderate level of variation, ranging from a minimum of 342 to a maximum of 1,062, with a standard deviation of 210. For the purpose of comparison, we also obtained a random sample of 13,768 customers who had yet to subscribe to the program as of July 2017. The data consist of two parts: transaction data and program usage data. The transaction data contain detailed information on each purchase made by a customer, when a customer purchased a product and how much she paid for it. The program usage data contain information on how a member benefited from the program, e.g., amount spent with gift cards, free samples received, etc. Our data also contain socio-demographic characteristics of customers, e.g., age, gender, and home address, which we utilize to control for customer heterogeneity while explaining the drivers of purchase behavior.

Using purchase data, we define a set of outcome measures associated with customer purchases. As the program is at the online channel only, unless specified otherwise, these measures are based on online purchases where the program could have a direct effect, and are constructed at the customer-month level, which is the unit of analysis in this research.[7] As our primary interest is to assess how effective the program is in lifting sales, our primary measure is the amount spent by a customer per month.[8] In addition, we consider two other (monthly) measures of customer purchases—number of purchases made (purchase frequency) and basket size ($) conditional on purchase (basket size)—because the change in purchase amount through subscription program can arise in multiple ways. For example, the program could lift purchase amount due to the increase in purchase frequency and/or basket size. Purchase frequency and basket size could also change in opposite directions, but the overall change in purchase amount might still be positive.

We also try to characterize the variety in purchase behavior with a few metrics. We classify a product (and its product category) a customer purchased as a new versus known product (category) on the basis of whether or not she had purchased it in the pre-subscription period. The first set of metrics relates to the variety at the product level: amount spent for new versus known products. The second set of metrics relates to the variety at the category level:

[7] As a robustness check, we investigate the impact of the program on total purchases from both the online and offline channels in Section 6.1.
[8] All transactions were recorded in the currency of the country in which the headquarters of the company was located. We converted purchase amount to U.S. dollars using the average exchange rate over the data period.

amount spent for new versus known categories.[9] As a proxy for engagement to the firm, these measures are useful to investigate how customers change their behavior in terms of the variety of new (versus known) products and categories after they joined the program.

4. Method

In this section, we first discuss the empirical strategy and treated and control groups to establish the causal effect of the subscription program. We next discuss the difference-in-differences approach and generalized random forests procedure followed by details about implementation.

4.1 Empirical Strategy

We wish to identify the causal impact of joining the program on subsequent purchase behavior. A key challenge for addressing this objective is due to self-selection — members may differ from non-members even before they joined the program and we would have biased results if we estimated the causal effect of subscription by directly comparing purchases between members and non-members. We also seek to decompose the total effect of the program on purchase behavior into that due to becoming a member per se (i.e., net effect) and that due to the economic benefits offered. Finally, we are interested in examining the heterogeneity in the treatment effect across customers. A typical way to accommodate heterogeneity is to slice data into subgroups based on covariates and investigate the effect in these subgroups. An issue with this approach is that it can become cumbersome as the covariate space increases and moderates the treatment effect in a non-linear manner.

[9] Skin care, make-up, hair care, bath and body care, fragrance, etc. are the main product categories of beauty products. Based on conversations with the retail partner, we decided to have five product categories for our empirical analysis in order to correspond to the way in which the firm monitors key metrics regarding customer purchases. They include skincare, make-up, hair care, bath and body care, and others in which we aggregated and grouped fragrance and the rest of the categories (e.g., tools and brushes).

To control for self-selection and identify the causal effect of subscription on purchase behavior at the individual level, we rely on a quasi-experimental design. Our baseline model uses a difference-in-differences (DD) approach (Angrist and Pischke 2008) and controls for selection based on any time-invariant unobservables. We also control for the marketing mix that a member was exposed to and quantify the net effect of membership on purchase behavior. We complete our modeling framework by embedding the DD specification within a generalized random forests procedure (Athey et al. 2019). Briefly, we estimate the DD specification for each member using a subsample of comparable customers defined by a random forest in a high-dimension covariate space. In doing so, we account for selection based on observables and heterogeneity across customers in a non-parametric manner and obtain individual-level treatment effects.

4.2 Treated and Control Groups

We focus on a cohort of members who joined the program around the same time in our main analysis. Such a cohort-level analysis is common when analyzing customer value (e.g., McCarthy et al. 2017). Focusing on a cohort of members is conducive to examining the causal effect of the program, as it gives well-defined pre- and post-treatment periods for the analysis. Our main findings consider the cohort of 721 members who joined the program in April 2016, four months after the launch of the program.[10] As reported in the robustness checks in Section 6, our findings are robust when we consider the effect of the program among members who joined in other months.

Before we establish the causal effect of the program on customer purchases, we examine the purchase amount for members and non-members over a 24-month period: April 2015 to March 2017. Of these, the first 12 months (i.e., April 2015 to March 2016) are prior to their

[10] In order to mitigate the concerns for selection and unmeasured confounders, we deliberately excluded early adopters as they may systematically differ from other customers (e.g., Rogers 2003).

subscription. Figure 1 offers model-free evidence that purchase behavior differed considerably between members and non-members, which also persisted over time. On average, members spent \$43.16 per month post subscription while non-members spent only \$3.93.

Insert Figure 1 about here

To assess whether non-members were similar to members before joining the program, we compare them on their purchases during the 12-month period prior to the subscription and their individual characteristics. Table 1 shows that members and non-members differed significantly on both their purchases and demographics. On average, members spent more per month than non-members (diff. = 5.61, $p < 0.001$), which is consistent with the intuition that customers who spent more were more likely to join the program as they could benefit more from the program. Members were older than non-members (diff. = 3.10, $p < 0.001$). Clearly, estimating the effect by merely comparing customer purchases between the two groups will be biased.

Insert Table 1 about here

One improvement on the naïve cross-sectional comparison is to construct a DD estimator by comparing the change in customer purchases over time for the members with that for the non-members. In this way, any time-invariant heterogeneity could be removed. In the next section, we lay out our DD specification. In Section 4.4, we discuss the validity of the identifying assumption of DD in our context and a machine learning approach to facilitate identification.

4.3 Difference-in-Differences

The difference-in-differences method attempts to control for time-invariant unobserved variables that bias estimates of the causal effect. Members of a subscription program may change their behavior due to membership and/or economic benefits offered to them. To measure the net impact of membership separately from the total effect of the program, we further control for the marketing mix that a member was exposed to. We estimate the following DD model:

$$Y_{it} = \tau Member_{it} + Z_{it}\beta + \alpha_i + \gamma_t + \epsilon_{it}, \tag{1}$$

where Y_{it} is the outcome measure, such as purchase amount, of customer i in month t. The indicator variable $Member_{it}$ for membership status equals one if customer i was a member in month t and zero otherwise. The two parameters α_i and γ_t are customer- and month-level fixed effects, respectively, and ϵ_{it} is the error term. By including the two-way fixed effects, we control for time-invariant customer characteristics as well as common time trends and month-to-month fluctuations. The vector Z_{it} contains any marketing mix customer i was exposed to in month t and the vector β contains the respective sensitivity parameters. Of primary interest is the parameter τ, which estimates the net effect of the program on purchase behavior, after controlling for marketing mix. If marketing mix were not separately accounted for in Equation (1), the parameter τ would then capture the total effect of the program on purchase behavior. In Section 5, we present both the total and net treatment effects on purchase behavior.

In our context, members receive unlimited free shipping service and member-exclusive benefits. We observe the impact of free shipping and samples is quite limited.[11] In contrast, we observe members benefited from price discounts and gift cards substantially: Members had an average 6% additional discount as compared to non-members and about 20% of monthly gift cards were redeemed. Therefore, we consider these two marketing mix variables (Z_{it}) to measure the net impact of the program on customer purchases.

[11] The online website offered all customers, regardless of the subscription, free shipping on orders above a certain threshold, which was satisfied by most orders. We also find less than 1% of purchases were induced by free samples, when we examined whether a customer purchased a product after receiving a free sample of that product.

While program usage data suggests the change in purchase behavior is likely related to program benefits among members, associating the change in purchase to benefits utilized is problematic due to reverse causality. For instance, frequent redemption of gift cards could be an effect of a high level of purchase rather than its cause. Next, we discuss how we operationalize two marketing mix variables to alleviate such issues.

We first construct a store-level price variable for customer i in month t as the weighted average price of a basket of products following the literature (e.g., Dube et al. 2018). There are two notable aspects of our operationalization of the store-level price variable. First, we observe the price paid for a product by a customer upon purchase. But we do not observe all the prices and discounts that a customer was exposed to while shopping on the website. As our goal is to characterize subscription-related prices paid by members as compared to non-members, we calculate a store-level price variable for customer i in month t as the weighted average price of a basket of products and allow it to vary at the group level (members and non-members) on a monthly basis.[12] Second, we fix the basket of products and weights to be common between the two groups to ensure the resulting store-level price captures different prices paid by the two groups rather than their differing product choices. We assume the monthly basket of products was comprised of those purchased by both members and non-members in that month. We select the frequency of a product purchased by members as the weight for a given product. Our group- and store-level price measure provides appropriate information on the level of prices that the members were exposed to due to the subscription program and how it varied on a monthly basis. Please see Web Appendix A for more details.

[12] We chose the minimum price paid for a product by anyone in a group in that month as the monthly price for the product. In Web Appendix A, we investigate a few alternative ways to construct the store-level price measure.

To avoid the reverse causality concern discussed earlier, we also include the amount of gift cards offered instead of the amount that customers used each month. However, all members received a $3 gift card each month, which is then confounded by membership status. To separately identify the net effect of the program from that of economic benefits, we make two assumptions: Gift cards affected purchase behavior by increasing the amount of spend that a customer allocated to the focal products in this study, and the amount of spend allocated, in turn, had a diminishing marginal return on customer purchases on the focal website (Hausman et al. 1994). We operationalize the baseline spend of a customer (without gift cards) by using her maximum monthly spend one year prior to subscription. We use a functional form with natural log to reflect both assumptions. Put together, Equation (1) can be written as follows:

$$Y_{it} = \tau Member_{it} + \beta_1 \log(Price_{it}) + \beta_2 \log(Baseline_i + Giftcard_{it}) + \alpha_i + \gamma_t + \epsilon_{it}, \quad (2)$$

where $Price_{it}$ is the store-level price encountered by customer i in month t, $Baseline_i$ is the baseline spend of customer i, and $Giftcard_{it}$ is the amount of gift card offered to customer i in month t. The parameters β_1 *and* β_2 are the semi-elasticity of price and gift card, respectively.

We exploit the panel structure of our data and heterogeneity in the baseline spend. Our panel consists of data over a 24-month period, where the first 12 months are prior to subscription. Based on purchase patterns shown in Figure 1, we excluded customer purchases in the month of adoption (April 2016) to avoid any potential simultaneity bias with the adoption itself. We use the month-to-month variation in purchase amount and store-level price to identify the price coefficient (β_1). Once β_1 is identified, parameters β_2 and τ are identified based on the functional form. We assume gift cards had a diminishing marginal return depending on the baseline spend in a logarithmic form. The parameter β_2 is identified by the difference of the unexplained increase in purchase amount upon subscription across members who differ in terms of the baseline spend. For example, a positive β_2 indicates the increase in customer purchases declines as the baseline spend increases. The parameter τ is identified by the remaining change in member purchases upon subscription and provides an estimate for the impact of membership were a customer to join and receive no economic benefits. Web Appendix B describes how different patterns in purchase data could help identify the model parameters.

4.4 Generalized Random Forests

While the DD model controls for time-invariant heterogeneity and estimates the average treatment effect, there are two issues worth discussing. First, note from Figure 1 and Table 1 that the difference in purchase amount between the two groups widens over time, suggesting that compared to an average non-member, the monthly spend by members increased over time even before their subscriptions. Thus, the parallel time trend assumption may not hold and the validity of the DD estimator is questionable (e.g., Bertrand et al. 2004).[13] Second, a typical way to accommodate heterogeneity in the treatment effect is to interact the treatment dummy with individual characteristics. This approach can become cumbersome as the covariate space increases and moderates the treatment effect in a non-linear manner.

Recent developments in the machine learning literature allow us to address both the above issues in a principled manner. We employ the generalized random forests (GRF) method (Athey et al. 2019) in this research. Similar to other methods for causal inference using observational data, e.g., kernel matching (e.g., Hastie et al. 2009), propensity score matching

[13] We tested whether purchase trends were common prior to subscription. To that end, we extended the DD specification in Equation (2) and estimated the difference in purchase behavior between members and non-members in month m using the following specification: $Y_{it} = \sum_m \tau_m \cdot I(t = m) \cdot Member_i + Z_{it}\beta + \alpha_i + \gamma_t + \epsilon_{it}$, where the indicator variables $I(t = m)$ are 1 if month t is m and $Member_i$ indicates whether customer i belongs to the treatment or control group. We normalized the first month in the pre-treatment period as the baseline of 0. Thus, the parameter τ_m captures the average difference in purchase measures between members and non-members relative to the baseline. We find the estimates are positive and statistically significant in the months close to the subscription, e.g., December 2015 to March 2016. We also estimated the model without marketing mix, and again find the parallel trend assumption does not hold. Results are available from the authors upon request.

(e.g., Hirano et al. 2003), and synthetic control (e.g., Abadie et al. 2010), the key idea for random forests is to define for each member a weighted set of neighbors that shares similar covariates and fit the model of interest using these neighbors. As an improvement to the traditional methods where the weights are chosen by deterministic kernel functions (kernel matching), parametric models (propensity score matching) or trend matching (synthetic control), Athey et al. (2019) propose to learn the weights using a revised random forest algorithm that is designed to minimize the estimation error.

Given the forest, we can define for each member with covariates x a weighted set of her neighboring customers by locating which customers fall into the leaves that contain the same covariates and associated frequency. The treatment effect for this member is estimated by fitting the DD specification on the weighted set. As compared to other commonly used methods, GRF is non-parametric and robust to model misspecification. The tree structure and the ensemble of many trees naturally account for complex interactions among covariates. The adaptive nature in trees can substantially increase the accuracy of the weighting function with a large space of covariates. Another advantage of GRF is that it uncovers the point estimates and confidence interval of the treatment effect at the individual level with formal asymptotic guarantees. These estimates allow us to explore heterogeneous treatment effects in a systematic manner and can sharpen our understanding of underlying drivers for the success of the program.

4.5 Implementation

For the outcome variables, we analyze transaction data over a 24-month (April 2015 to March 2017) period because we are interested in examining the causal effect for the long term. Of these, the first 12 months are prior to subscription. As noted earlier, we excluded customer purchases in the month of subscription adoption (April 2016).[14]

To control for potential confoundedness, we include three sets of variables that describe members and non-members in the pre-treatment period. The first set of covariates relates to the customer-firm relationship, which would be associated with the adoption of a service, namely, tenure, breadth, and depth (e.g., Bolton et al. 2004, Prins and Verhoef 2007). We calculate tenure based on elapsed time since having an account on the website. We measure breadth by the number of unique categories purchased and depth by the number of transactions made. We also include the average basket size. In addition, we include monthly purchase amount per category during the 12-month pre-subscription period, instead of the total amount across product categories, because it could help find clusters of customers with similar purchase patterns across product categories. We also include the variance of monthly purchase amount because it could relate to customer response for unlimited free shipping service.

The second set of covariates relates to psychographic measures that reflect personality traits, preferences or interests, values and attitudes (e.g., Baumgartner 2002). As psychographics are not directly observable, their measures are more nuanced. Given that we use observational data with no surveys sent to customers, we explore a few measures associated with customer purchases by summarizing certain aspects of purchase behavior: exploratory (purchases of new products), repetitive (repeat purchases of a product), and promotional (discounts received for purchases).[15]

[14] The estimate we report provides a conservative estimate of the effect on customer purchases. Detailed results, which include April 2016, are available from the authors upon request.

[15] We conducted our analysis with alternative operationalization of the variables. For instance, we estimated the model with an alternative measure of repetitive purchase, e.g., share of repeat product purchases per month and category. We find the results are largely similar.

The third set of covariates relates to socio-demographics of customers. We include age and gender. We also include the coordinates of a customer's home address because it can help control for other unobserved socio-demographics that affect subscription, e.g., education, income, life style, and so on. Table 2 summarizes the covariates and describes how the variables are operationalized. Altogether, we use 72 covariates to build the trees for the random forest.[16]

Insert Table 2 about here

We refer readers to Athey et al. (2019) and provide the implementation details in Appendix A. We briefly describe the procedure for building the forest. To build the random forest, we first grow a decision tree by iteratively partitioning the data into subgroups (leaves). In each iteration, the algorithm chooses the covariate and the cutoff to find partitions where treatment effects most differ. As a single tree is likely to overfit the data, an ensemble of trees is generated. For each tree, a random set of the covariates can be potentially used to form the splits. The number of trees, minimum number of treatment and control observations in each leaf, subsample size, and size of the set of covariates used to build each tree are the hyper-parameters of a forest, which we choose by cross-validation. In the context of panel data, we also account for clusters at the individual level in the sampling as well as estimation process.

The forest performs well in balancing the members and non-members. Following Imbens and Rubin (2015), we use normalized absolute mean difference to assess the degree of balance of the observables. Figure 2 shows the normalized mean difference of the variables before and after the adjustments by GRF. After the adjustment, the normalized absolute mean difference of the

[16] Before we build GRF for matching, as an important first step to ensure the final estimates are credible and robust, we pre-processed data by excluding customers with extreme propensity to adopt the program and improved overlap in covariate distributions (Imbens and Rubin 2015). We estimated the propensity scores by predicting membership with the covariates using a regression forest. We find 50 of the members have propensity scores with no counterparts in the control group. We thus excluded those customers whose propensity scores do not lie on the common support. Our final estimation was conducted on a sample of 671 members and 11,745 non-members.

variables between the members and the matched non-members are mostly below 0.1. And the members and non-members are indistinguishable in terms of their observed characteristics. We also test whether parallel time trends hold in our context using a series of placebo tests. We define a placebo treatment on members six months prior to when the actual subscription took place and estimate the effect of such a placebo treatment on the outcome variables of interest using a DD model on all the weighted pre-subscription data. If the parallel trend assumption holds, we should expect null effects. We find along all dimensions of interest, there is no change in member purchases six months prior to the subscription.[17] These results indicate that the members and matched non-members have statistically comparable pre-treatment time trend.

Insert Figure 2 about here

Finally, given the forest, we define for each member a weighted set of her neighboring customers, which may include both non-members and other members. For a member with covariates x, the weights are the frequency with which each customer falls into the leaves that contain x.[18] The procedure then fits Equation (3) with the weighted set of observations for each member. Hence, we effectively combine GRF with DD specification after controlling for economic benefits in the subscription program. To obtain the treatment effect across members, we construct a doubly robust average treatment effect estimator by augmenting the naïve plugin estimator with a residual-based correction. The doubly robust estimator combines results from GRF and a regression-based prediction such that it is robust to mis-specification of either the matching model or regression model (Chernozhukov et al. 2018).[19]

[17] Web Appendix C reports detailed results on the placebo tests.
[18] We use out-of-bag predictions to avoid over-fitting, i.e., only trees that do not include the member during tree building are used to produce the weights.
[19] Web Appendix D describes how the individual-level estimates are identified.

5. Findings

In this section, we first discuss the main findings on the treatment effect on customer purchases. We document both types of effect of the program, namely, total effect of the program on customer behavior and net effect for the impact of membership to the program after accounting for marketing mix. We also discuss how the effects (both total and net) vary over time and across customers. Finally, we explore some possible explanations underlying the effect.

5.1 Average Treatment Effects

5.1.1 Total Treatment Effects

In Table 3a, we report the average total treatment effect on customer purchases.[20] Because our primary objective is to identify the total effect for the long term, we first discuss the estimates over a 12-month period after joining the program. The first column in Table 3a shows that on average, purchase amount per month among members increased by $27.45. The total effect of the program is economically significant, as average purchase amount per month was about $12 before subscription.

Insert Table 3a about here

The treatment effect on purchase amount is quite striking. As shown in Figure 1, however, patterns in our data suggest that our finding is not an artifact. We conducted a comprehensive review of extant literature and observe the impact of membership programs on customer behavior has substantial variation, ranging from no effect at all to as high as 150% increase in customer response across various empirical settings.[21] For instance, using an

[20] Web Appendix E reports the results in which the outcome measure was transformed with natural log in the estimation.

[21] Lal and Bell (2003) examine the impact of frequent shopper programs in grocery retailing and find an increase of $98, $141, and $150 across three different segments, respectively. As they do not report the baseline prior to the frequent shopper program in their study, however, we are not able to compute the relative impact of the program.

application of “a buy ten get one free program” offered by a golf course, Hartmann and Viard (2008) find no changes in customer response. Similarly, Lewis (2004) evaluates a loyalty program from an online merchant and finds about 2% increase in customer revenues. Kopalle et al. (2012) use data from the loyalty program of a hotel chain, and find about a 30% increase in customer spending due to the program. Specifically, the price-oriented segment increases its spending by 29% (from $343.7 to $443.8), and the service-oriented segment increases its spending by 34% (from $1189.3 to $1589.6). Using data from a convenience store chain’s loyalty program, Liu (2007) finds consumers whose initial patronage levels were low or moderate considerably increase their spending by around 150% under the loyalty program.

We note that while both Hartmann and Viard (2008) and Kopalle et al. (2012) study customer response to reward programs in a single category (e.g., golf course, hotel), Liu (2007) examines the impact of a loyalty program on customer behavior when the firm offers multiple different categories (e.g., household and grocery, personal and health care, beverage and food, etc. at a convenience store). Our empirical context is close to Liu’s because the focal firm offers a wide assortment of brands and products across different categories, ranging from low-end to high-end goods at varying prices. We believe that the effect size in our study is an important addition to the literature because our study examines customer response in a contractual subscription program where a one-time purchase of a subscription can lead to recurring sales. Subscribers are more likely to make repeat purchases across categories during their contract in the subscription. In contrast, existing literature examines customer response in non-contractual reward programs.

The increase in customer purchases could be driven by the increase in purchase frequency and/or basket size. These two metrics, although typically positively correlated, could have different implications for the firm. Our results show members made about one additional purchase per month (1.15) post subscription. Interestingly, we find a small but significant decrease in the basket size ($-5.14). Recall that in our context the threshold for free shipping was relatively low for non-subscribers and the value of free shipping service was quite limited. One possibility is that members might make their basket into smaller ones to utilize, for example, recurring monthly gift cards. Our results on the net effects of the subscription later confirm this intuition.

We next examine the effect of subscription on the variety in purchase behavior. Recall that we classify products (categories) a customer purchased to new versus known products (categories) based on prior purchase behavior. We find a significant increase across all variety measures.[22] At the category level, more than 95% of the increase in purchase amount ($26.19 out of $27.45) came from known categories, and the rest from new categories that customers had never purchased from prior to subscription. At the product level, approximately 75% of the increase in purchase amount ($20.89 out of $27.45) was from new products that a customer had never purchased. Taken together, our evidence supports that subscription makes members purchase more frequently, with a greater variety of products and categories, leading to increased customer loyalty and share of wallet.

Additionally, we examine the temporal variation in the treatment effect on purchase behavior. For instance, the program could create an initial excitement among members, leading to increased purchase. If the novelty effect of the program were the only underlying reason for

[22] We note that the treatment effect on purchase amount is slightly different from the sum of the treatment effect on purchase amount in known versus new products and categories. Because our matching procedure aims at minimizing the mean squared error, matches depend on both the covariates and outcome variable. Therefore, matches can differ slightly when evaluating the treatment effect on different outcome variables. However, unconfoundedness guarantees that all the estimates are unbiased.

the behavioral change, the effect will likely fade away over time and the program would have limited impact on the firm's long-term revenue (e.g., Galak and Redden 2018). To examine the temporal effect of the program on purchase amount, we utilize the purchase patterns shown in Figure 1 and distinguish between treatment effects within the first two months (excluding the first month of adoption), the next two months, and the remaining months in the post-subscription period. We estimate the temporal effects by applying GRF on data in the corresponding time periods relative to the 12-month pre-treatment period. We find the effect on customer purchases is the largest (an increase of $40.57) within the first two months and persists (an increase of $25.53 per month) after four months upon joining.

In summary, there is a causal impact of subscription program on customer purchases. The total effect of the program on customer purchases is economically and managerially significant and persists over time. The program keeps customers more engaged both in terms of frequency and variety in their purchases.

5.1.2 Net Treatment Effects

We next discuss the net effect of becoming a member per se after marketing mix is controlled for. Table 3b reports the average net treatment effect of the program. Interestingly, we find only one-third of the total effect on purchase amount is due to the economic benefits and the remaining two-thirds ($17.91 out of $27.45) is attributed to becoming a member.[23]

[23] We calculate the relative contribution of price discounts and gift cards, in which the change in purchase behavior explained by marketing mix is obtained by multiplying the change in its level and sensitivity to marketing mix. Our results suggest price discounts accounted for about $4.20 increase in purchases and $3 gift cards explained about $5.30 increase in purchases. And the price and coupon elasticities are qualitatively similar to those documented in past research and have the face validity. Specifically, the implied price elasticity (the percentage change in purchase amount in response to a one percent change in price) is -1.75 and the implied marginal effect of coupon (the dollar change in purchase amount in response to a one-dollar change in coupon) is 1.76. As a point of comparison, in two meta-analyses, Tellis (1988) and Bijmolt et al. (2005) report average price elasticities of -1.76 and -2.62, respectively. And Venkatesan and Farris (2012) report the marginal effect of coupons to be greater than 2, as the mere exposure to coupons can help lift revenue.

Insert Table 3b about here

Looking at the temporal patterns, we find the increase in purchase amount due to becoming a member is largest (an increase of $37.28) within the first two months post subscription. While it gradually faded away over time, the effect is persistent and managerially important. Across most of the other metrics as well, a significant part of the increase in customer purchases was not accounted for by the economic benefits members were exposed to. The net effect on basket size is not significant, indicating that after accounting for economic benefits, there is no further impact on the basket size. In sum, even after controlling for the economic benefits of the program, a significant part of the impact of the program on customer purchases can be attributed to becoming a member per se.

5.2 Heterogeneous Treatment Effects

Figures 3a and 3b show the distribution of the total and net treatment effects on purchase amount, respectively. Both figures present that the impact of subscription on customer purchases is heterogeneous across customers. These results illustrate the benefits of employing GRF, in that we can obtain individual-level treatment effects.

Insert Figures 3a and 3b about here

Figure 3a shows significant variation in the total effect across members, ranging from less than $10 to around $80. About 20% of the members increased their purchases by $15 or less, and approximately 27% increased their purchases by $40 or more. Figure 3b illustrates that there is large heterogeneity in the net effect as well. About 40% of the members increased their purchases by $15 or less, and around 14% increased by $40 or more. The proportion of the total effect explained by simply becoming a member also has substantial heterogeneity across members. On average, 65% of the total increase in purchase amount was driven by membership.

For 15% of members, almost 90% of their total increase in purchase amount could be attributed to the economic benefits of the program. In contrast, 40% of members would change their behavior even without economic benefits, in that more that 80% of their total increase in purchase was due to just becoming a member of the program. These results show that paid customer programs (e.g., subscription, reward) can have an impact on purchase behavior even after accounting for their economic benefits (e.g., Bolton et al. 2000).

5.3 Possible Explanations

We propose four underlying drivers based on extant literature. The four drivers each generate unique predictions regarding the temporal purchase patterns. Hence, to facilitate a sharp test of whether these drivers are at work, we obtain more granular estimates of member-specific treatment effects at the monthly level following the estimation procedure proposed in Section 4. Based on these member- and month-level estimates, we use a unified panel regression to examine the driver(s) at work. We caution that with only observational data it is challenging to conclusively point to a single explanation, as multiple drivers can be at work.

Our first driver is based on a unique feature of subscription programs. Unlike members of (free) loyalty programs, members of a subscription program pay an initial fee to join the program. While rational customers should not take this cost into account when making subsequent purchase decisions, extant work suggests people exhibit sunk cost fallacy and tend to increase their purchases to justify their initial upfront payment (e.g., Thaler 1980, Arkes and Blumer 1985). We hypothesize that sunk cost can be especially salient for subscription programs and induce an increase in purchase amount beyond what the economic effects of the program could explain.

We also investigate the presence of other drivers based on extant research on loyalty programs (e.g., Bolton et al. 2004). A membership program can induce positive affect towards the firm and lead to increased purchases (e.g., Leenheer et al. 2007). Members may also feel superior to other customers when they have access to exclusive offers, and their enhanced status can encourage purchases as well (e.g., Drèze and Nunes 2009). Additionally, a program may encourage customers to purchase upon joining, and this increase in the short run may lead to a habitual increase in the long run (e.g., Wood and Neal 2009). While each driver is predicted to increase subsequent purchases, there are differences amongst them with respect to how some program benefits moderate their impact. Next, we discuss this aspect for each of the four drivers introduced above.

Past research on sunk cost fallacy suggests that as customers accumulate the use of a product or service after an upfront payment, sunk cost becomes less salient and its effect wears off (Ho et al. 2017). Thus, if members responded to sunk cost, their increase in purchase (after controlling for economic benefits) should decrease with how much they benefited from the program.

A prediction based on positive affect is that customers act favorably towards the program shortly after they become a member. However, as they continue to purchase from the firm, they will experience hedonic decline (e.g., Galak and Redden 2018). Comparing the purchase pattern based on positive affect with that based on sunk cost fallacy, their temporal trends may appear similar. However, the monthly program benefits (related with sunk cost) may vary independently from monthly purchase (related with positive affect) as the level of discounts vary from one month to the next.

The presence of the third driver, status, would suggest that the effect of the program on purchase would diminish as membership becomes less exclusive. Extant research on membership programs suggests that the value of status created by membership is associated with its distinctiveness (e.g., Grier and Deshpandé 2001). If members derive status from the program, we would expect that the program should have a smaller impact on their purchases as the number of members increase.

The fourth driver, habit formation, would predict that customers' high level of purchase in the long term is a result of state dependence (i.e., habits formed) based on the increase in purchase in the short term. If habit formation were the only mechanism at work, when comparing the purchase behavior of members with that of non-members that have similar purchases not only before subscription but also shortly after subscription, we should not expect any residual difference in their future purchases. If, however, there are still differences in future purchases of members (versus non-members) after controlling for this comparison, they must stem from reasons other than their consumption habit.

In summary, the four potential underlying drivers predict differing temporal patterns for the net effect of the program and purchase behavior, allowing us to investigate the mechanism at work.

We propose a unified framework to investigate the impact of all four underlying drivers on purchase behavior. Our framework leverages the individual- and month-level treatment effects from GRF. Under the framework, we first tease out the effect of habit formation by comparing members and non-members with similar purchase patterns from 12 months before subscription to a few months after subscription (we choose a four-month period post subscription in line with our analysis for temporal effects). The remaining treatment effect of the program from this matched sample is sans any habit formation after subscription (assuming that the first few months post subscription are critical for forming consumption habits). These remaining treatment effects then become the dependent measure in a panel regression where we investigate their association with the moderators to test the presence of the other three drivers. In what follows, we describe each step in more detail.

We determine the average treatment effect of the program (sans habit formation) on customer behavior after four months of subscription. Specifically, we build a GRF using all covariates in the main analysis with data from April 2015 (12 months before subscription) to July 2016 (four months after subscription) to estimate the changes in monthly purchases after July 2016 relative to the pre-subscription period.[24] We find that, on average, members increased their purchases by \$11.91 per month after four months of subscription (as compared to \$12.31 in Table 3b, which is inclusive of habit formation). This comparison suggests that habit formation is not a big driver of the net effect.

In order to study the temporal patterns of the net treatment effect, we obtain more granular estimates of member-specific treatment effects at the monthly level following the same estimation procedure (i.e., the impact of habit formation is already accounted for). For member i and month t, let $\hat{\tau}_{it}^{T}$ and $\hat{\tau}_{it}^{N}$ denote the member- and month-level total and net treatment effect, respectively, after accounting for habit formation. We estimate the following specification:

$$\hat{\tau}_{it}^{N} = \gamma_1 \cdot CumBenefit_{it} + \gamma_2 \cdot CumPurchase_{it} + \gamma_3 \cdot CumPurchase_{it}^{2} + \gamma_4 \cdot NMember_t + \delta_i + \xi_{it}, \quad (3)$$

[24]We also replicated the analysis using an alternative time period of six months post subscription. We find the qualitatively similar results. We thank a reviewer for suggesting this robustness check.

where the variable $CumBenefit_{it} = \sum_{s<t} \hat{\tau}_{is}^{T} - \hat{\tau}_{is}^{N}$ captures the cumulative impact of economic benefits from the program for member i prior to month t.[25] The variable $CumPurchase_{it}$ is the cumulative purchase amount for member i prior to month t. We include the quadratic term as well to capture the different possible patterns of hedonic decline (Galak and Redden 2018). The variable $NMember_t$ is the total number of members who had adopted the program till month t. We include individual-fixed effects (δ_i) and ξ_{it} is the error term Note that as we include the cumulative number of adopters over time, we cannot separately estimate time-fixed effects. As the dependent variable in the above regression is estimated with error, we use generalized least squares to estimate the model to account for the variance of the dependent variables (Hanushek 1974).

Table 4 reports results from the panel regression. The first column presents the results from a regression where the cumulative benefit of the program is the only explanatory variable. The results indicate that consistent with sunk cost fallacy, the increase in purchase was negatively correlated with the program benefits. Column (2) reports the results with a linear and a quadratic term for cumulative purchase also added in the regression. The results support the presence of hedonic decline with past purchase. Column (3) adds the number of members as an additional explanatory variable and we find the number of members also significantly affected the net effect. Taking these results together, we find a higher net effect is associated with lower program usage, lower past purchase, and fewer members. Recall that we have already accounted for state dependence in the treatment effect. While we cannot interpret these results as causal, the pattern of results is consistent with the hypotheses that members increase their purchase due to

[25] Since our price variable is at the group level and ignores the within-group differences, the estimated economic effects might be measured with error at the individual level. As long as this measurement error is independent over time, i.e., each customer's true price fluctuates around the group-level price, it should not bias our results.

positive affect towards the firm and status effect. Importantly, after accounting for all the above patterns, we find evidence that customers exhibit a sunk cost fallacy.

Insert Table 4 about there

6. Robustness Checks

In this section we analyze the robustness of our findings by using alternate outcome measures and treated groups and addressing potential effects of unobservables.[26]

6.1 Alternate Outcomes

As the subscription program we study is an online-only program, our main analysis focused on customer purchases only on the website. The firm we partnered with has both a brick-and-mortar and online presence and is able to link customer purchases between online and offline channels at the individual level through its reward program. We thus investigate whether the increase in online purchases through subscription was due to the channel-switching behavior to online from offline (e.g., Forman et al. 2009, Wang and Goldfarb 2017).

We perform the analysis described in Section 4 by replacing online purchases with total purchases combined between online and offline channels and retaining the operationalization of all covariates in Table 2. We find the treatment effect on purchase amount is $26.72 (std. err. = 1.68), suggesting that in our context, the online and offline channels are only weak substitutes.[27] In sum, the program is effective in lifting overall revenue for the firm.

[26] We estimated both the total and net effects of the program on all outcome measures across different time windows as we did in our main analysis. As our primary interest is on the long-term effect on purchase amount, we only reported the total effect on purchase amount based on all months post subscription. Other results are available from the authors upon request.

[27] We also performed the analysis by further controlling for additional covariates using offline purchases, e.g., monthly purchase amount at the category level. We find the results are similar.

6.2 Alternate Treated Groups

In our main analysis, we used a single cohort of the members who joined the subscription program in April 2016. As a robustness check, we replicate the analysis for members who joined the program during other months. Our results, as shown in Table 5, suggest that the treatment effect on purchase amount across several cohorts is largely similar and our results are robust.

Inset Table 5 about here

6.3 Selection on Unobservables

The GRF framework has advantages over other causal inference methods, e.g., propensity score matching and nearest neighbor matching, in that it matches members and non-members in a non-parametric and robust manner. Since the treatment is not assigned randomly, the validity of the method still hinges on the assumption of unconfoundedness. Although we used a rich set of covariates based on individual-level data in our main analysis, the treatment status is independent of the potential outcomes. While the unconfoundedness assumption is usually not directly testable, we present an additional piece of evidence to alleviate the concern for this assumption.

We estimate the treatment effect using the members only. Specifically, we use late adopters, rather than non-members, as controls for early adopters. The late adopters could be better controls for the early adopters than non-members if their adoption time is close enough (e.g., Goldfarb and Tucker 2011, Manchanda et al. 2015, Datta et al. 2017, Narang and Shankar 2019). We choose customers who joined the program between August 2016 and November 2016 as the control group, allowing us to have enough customers to match from and enough time periods to estimate the treatment effect. We find qualitatively similar results. On average, customers increased monthly purchase amount by $31.38 (std. err. = 2.18).

7. Conclusions

In this paper, we study the causal impact of customers joining a subscription program on their subsequent purchase behavior. Utilizing a panel data at a company that launched a subscription program on its website, we measure the causal effect of the program on customer behavior. We combine the difference-in-differences approach with the generalized random forests procedure and obtain the treatment effect estimates at the individual level. We find subscription is effective in lifting sales and keeps members more engaged in terms of frequency and variety in their purchases. Interestingly, only one-third of the treatment effect on purchase amount is attributed to the economic benefits of the program. Both the total and net effects of the program on customer purchases are economically significant, persistent over time, and heterogeneous across customers. Our main findings are robust to different outcomes, different samples of the treated customers, and potential confounding effects from unobservables.

To uncover the underlying mechanism that leads to the behavioral changes, we leverage the individual- and month-level treatment effects in our estimation. In addition to multiple drivers (e.g., habit, status, affect) documented in the context of other types of membership programs (e.g., loyalty programs), notably we document evidence that members experience a sunk cost fallacy due to the upfront payment that subscription programs entail.

Our results shed light on the practice of subscription-based businesses for customer retention and development. As companies are increasingly concerned about customer engagement (e.g., breadth and depth of purchases in products and categories), it is important to understand how subscription programs affect future sales. In our context, we find such programs are broadly effective in lifting sales especially by making customers more frequent shoppers not only in the known categories but also in new categories. While we study one subscription program launched by a single firm, the insight that the effect of the program on customer purchases goes beyond the economic benefits is not limited to the structure of our focal program. Our results also enable us to derive recommendations related to the design of subscription programs. For example, given the sensitivity to sunk costs, it may benefit companies to make the initial payment more salient after customers become members. Our suggestion is in line with past work showing that making prices salient can make members consume a service on a more consistent basis (e.g., Gourville and Soman 2002). Our analyses also contribute to the ongoing debate concerning the amplified friction in the retail industry due to the rise of subscription programs (e.g., Amazon Prime). Our results suggest that firms are able to lock in customers with subscription programs by creating a sunk cost fallacy, which may potentially lead to increased market concentration.

As our research is among the first to identify the causal effect of subscriptions in the retail area on customer behavior, naturally there are limitations that should be acknowledged and addressed in future research. First, we have studied the causal effect of a subscription program in a given firm and it is likely that some of our findings could reflect the customer base and product categories of our partner firm. The subscription period is also reasonably long (one year) and early termination was not allowed, so sunk cost is prominent. In line with these notions, while the subscription program examined in this research is an online-only program, a few other programs in which members benefits from both online and offline channels (e.g., Amazon, Barnes & Noble). With that in mind, we hope our approach provides a framework for further studies in other product categories (e.g., consumer software, food preparation), on programs with various pricing schemes (e.g., programs with more varying time windows for subscriptions), and across channels in retail and other industries (e.g., financial services, business solutions). Second, our findings suggest that a subscription program creates value for the firm. In the long run, with the relevance and popularity of subscription-based businesses, it is possible that many more companies will have their own subscription programs. With competition in play, the effect of a subscription program on customer engagement and purchase remains unclear. One may expect that the effect of a new program may not be as large but there may also be some interactions among market and firm characteristics, features of subscription programs, and subsequent changes in customer behavior. We hope that our work will inspire further studies to deepen our understanding in this nascent and important area of research.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105 (490), 493-505.

Angrist, Joshua D. and Jorn-Steffen Pischke (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, NJ.

Arkes, Hal R. and Catherine Blumer (1985), "The Psychology of Sunk Cost," *Organizational Behavior and Human Decision Processes*, 35 (1), 124-140.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019), "Generalized Random Forests," *The Annals of Statistics*, 47 (2), 1148–1178.

Baumgartner, Hans (2002), "Toward a Personology of the Consumer," *Journal of Consumer Research*, 29 (2), 286-292.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119 (1), 249-275.

Bijmolt, Tammo H.A., Harald J. van Heerde, and Rik G.M. Pieters (2005), "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research*, 42 (2), 141–156.

Bolton, Ruth N., P. K. Kannan, and Matthew D. Bramlett (2000), "Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value," *Journal of the Academy of Marketing Science*, 28 (1), 95-108.

Bolton, Ruth N., Katherine N. Lemon, and Peter C. Verhoef (2004), "The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research," *Journal of the Academy of Marketing Science*, 32 (3), 271-292.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21 (1), C1–68.

Datta, Hannes, George Knox, and Bart J. Bronnenberg (2017), "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery," *Marketing Science*, 37 (1), 5-21.

Drèze, Xavier and Joseph C. Nunes (2009), "Feeling Superior: The Impact of Loyalty Program Structure on Consumers' Perceptions of Status," *Journal of Consumer Research*, 35 (6), 890-905.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi (2018), "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession," *Marketing Science*, 37 (1), 22–53.

Forman, Chris, Anindya Ghose, and Avi Goldfarb (2009), "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 55 (1), 47-57.

Galak, Jeff and Joseph P. Redden (2018), "The Properties and Antecedents of Hedonic Decline," *Annual Review of Psychology*, 69, 1-25.

Goldfarb, Avi and Catherine E. Tucker (2011), "Privacy Regulation and Online Advertising," *Management Science*, 57 (1), 57-71.

Gourville, John and Dilip Soman (2002), "Pricing and the Psychology of Consumption," *Harvard Business Review*, 80 (9), 90-96.

Grier, Sonya A. and Rohit Deshpandé (2001), "Social Dimensions of Consumer Distinctiveness: The Influence of Social Status on Group Identity and Advertising Persuasion," *Journal of Marketing Research*, 38 (2), 216-224.

Hanushek, Eric A. (1974), "Efficient Estimators for Regressing Regression Coefficients," *The American Statistician*, 28 (2), 66.

Hartmann, Wesley R. and V. Brian Viard (2008), "Do Frequency Reward Programs Create Switching Costs? A Dynamic Structural Analysis of Demand in a Reward Program," *Quantitative Marketing and Economics*, 6 (2), 109-137.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media.

Hausman, Jerry, Gregory Leonard, and J. Douglas Zona (1994), "Competitive Analysis with Differentiated Products," *Annals of Economics and Statistics*, (34), 143–57.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71 (4), 1161-1189.

Ho, Teck-Hua, I. P. L. Png, and Sadat Reza (2017), "Sunk Cost Fallacy in Driving the World's Costliest Cars," *Management Science*, 64 (4), 1761-1778.

Imbens, Guido W. and Donald B. Rubin (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Kopalle, Praveen K., Yacheng Sun, Scott A. Neslin, Baohong Sun, and Vanitha Swaminathan (2012), "The Joint Sales Impact of Frequency Reward and Customer Tier Components of Loyalty Programs," *Marketing Science*, 31 (2), 216-235.

Lal, Rajiv and David E. Bell (2003), "The Impact of Frequent Shopper Programs in Grocery Retailing," *Quantitative Marketing and Economics*, 1 (2), 179-202.

Leenheer, Jorna, Harald J. van Heerde, Tammo H.A. Bijmolt, and Ale Smidts (2007), "Do Loyalty Programs Really Enhance Behavioral Loyalty? An Empirical Analysis Accounting for Self-Selecting Members," *International Journal of Research in Marketing*, 24 (1), 31-47.

Lewis, Michael (2004), "The Influence of Loyalty Programs and Short-Term Promotions on Customer Retention," *Journal of Marketing Research*, 41 (3), 281-292.

Liu, Yuping (2007), "The Long-Term Impact of Loyalty Programs on Consumer Purchase Behavior and Loyalty," *Journal of Marketing*, 71 (4), 19-35.

Manchanda, Puneet, Grant Packard, and Adithya Pattabhiramaiah (2015), "Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community," *Marketing Science*, 34 (3), 367-387.

McCarthy, Daniel M., Peter S. Fader, and Bruce G.S. Hardie (2017), "Valuing Subscription-Based Businesses Using Publicly Disclosed Customer Data," *Journal of Marketing*, 81 (1), 17-35.

Narang, Unnati and Venkatesh Shankar (2019), "Mobile App Introduction and Online and Offline Purchases and Product Returns," *Marketing Science*, 38 (5), 756-772.

Prins, Remco and Peter C. Verhoef (2007), "Marketing Communication Drivers of Adoption Timing of a New E-Service among Existing Customers," *Journal of Marketing*, 71 (2), 169-183.

Raghubir, Priya (2004), "Free Gift with Purchase: Promoting or Discounting the Brand?," *Journal of Consumer Psychology*, 14 (1), 181-186.

Rogers, Everett M. (2003), *Diffusion of Innovations*, 5th edition, Simon and Schuster.

Tellis, Gerard J. (1988), "The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales," *Journal of Marketing Research*, 25 (4), 331.

Thaler, Richard (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1 (1), 39-60.

Venkatesan, Rajkumar and Paul W. Farris (2012), "Measuring and Managing Returns from Retailer-Customized Coupon Campaigns," *Journal of Marketing*, 76 (1), 76-94.

Wang, Kitty and Avi Goldfarb (2017), "Can Offline Stores Drive Online Sales?," *Journal of Marketing Research*, 54 (5), 706-719.

Wood, Wendy and David T. Neal (2009), "The Habitual Consumer," *Journal of Consumer Psychology*, 19 (4), 579-592.

Table 1: Summary Statistics of Members vs. Non-Members

Variable	Members	Non-Members	Difference	*p*-value
Purchase amount ($)				
04/2015	9.86	6.36	3.50	0.000
05/2015	9.77	6.68	3.09	0.000
06/2015	9.81	7.03	2.78	0.000
07/2015	9.54	5.84	3.70	0.000
08/2015	10.05	6.33	3.72	0.000
09/2015	11.17	7.25	3.92	0.000
10/2015	14.02	9.63	4.40	0.000
11/2015	11.18	7.27	3.90	0.000
12/2015	11.33	4.90	6.43	0.000
01/2016	16.08	5.50	10.57	0.000
02/2016	14.21	4.93	9.27	0.000
03/2016	16.81	4.77	12.05	0.000
Demographics				
Age	35.94	32.84	3.09	0.000
Gender (Female = 1)	0.93	0.94	-0.01	0.162
Observations	721	13,768		

Table 2: Covariates for Generalized Random Forests

Covariates	Operationalization	Mean	Std. Dev.
Customer-firm relationship			
Tenure	Elapsed time (year) since having online account	5.55	3.19
Breadth	Number of unique categories purchased	1.97	1.23
Depth	Number of transactions made	1.75	1.27
Basket size	Average basket size	33.51	31.26
Monthly purchase amount: Category-level	Monthly spend at each product category	1.33	5.91
Monthly purchase amount: Std. Dev.	Standard deviation of monthly purchase amount	15.07	20.77
Psychographics			
Exploratory	Inverse of average time (year) taken for the purchase among three new products purchased since the launch	3.12	13.39
Repetitive	1 if a customer made repeat purchases of a product more than ten times, 0 otherwise	0.03	0.18
Promotional	Average discount rate received for purchases	0.31	0.23
Socio-demographics			
Age		33.96	8.96
Gender	0 = Male, 1 = Female	0.94	0.23
Home address	Coordinates of home address		

Table 3a: Total Treatment Effects

	All months	Month 2	Months 3-4	Months 5+
Purchase amount ($)	27.45	40.57	28.95	25.53
	(1.75)	(4.70)	(2.33)	(2.03)
Purchase frequency	1.15	1.37	1.29	1.10
	(0.05)	(0.02)	(0.03)	(0.01)
Basket size	-5.14	-9.44	-5.01	-5.92
	(2.47)	(5.23)	(1.76)	(2.02)
Variety ($)				
Purchase amount in known categories	26.19	40.62	27.22	23.61
	(1.77)	(4.83)	(2.30)	(2.07)
Purchase amount in new categories	1.99	1.00	2.64	1.77
	(0.29)	(0.32)	(0.57)	(0.27)
Purchase amount of known products	7.13	6.02	7.65	7.18
	(0.68)	(3.05)	(0.84)	(0.73)
Purchase amount of new products	20.89	35.74	22.22	17.86
	(1.67)	(3.26)	(2.19)	(1.93)

Notes: Standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 3b: Net Treatment Effects

	All months	Month 2	Months 3-4	Months 5+
Purchase amount ($)	17.91	37.28	22.26	12.31
	(1.96)	(9.51)	(2.70)	(1.99)
Purchase frequency	0.97	1.20	1.24	0.80
	(0.04)	(0.11)	(0.07)	(0.05)
Basket size	-6.19	-2.88	-4.63	-6.67
	(12.77)	(8.89)	(2.80)	(10.47)
Variety ($)				
Purchase amount in known categories	17.44	35.74	21.77	12.36
	(1.96)	(7.22)	(3.00)	(2.46)
Purchase amount in new categories	1.40	1.06	1.44	1.27
	(0.23)	(1.24)	(0.58)	(0.16)
Purchase amount of known products	3.26	7.20	1.06	3.99
	(1.29)	(3.29)	(1.93)	(1.31)
Purchase amount of new products	14.23	28.59	20.47	10.56
	(1.42)	(6.02)	(2.19)	(2.02)

Notes: Standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 4: Regression Results

Construct	Variables	Net Effect (1)†	(2)†	(3)
Sunk cost	Cumulative benefit	-0.1195***	-0.1100***	-0.1001***
		(0.0054)	(0.0060)	(0.0068)
Affect	Cumulative purchase		0.0131***	-0.0091***
			(0.0026)	(0.0029)
	Cumulative purchase^2		-1.72e-06*	1.05e-06
			(8.99e-07)	(9.24e-07)
Status	Number of members			-0.0006***
				(0.0002)
	Individual-fixed effect	Y	Y	Y
	Observations	5,367	5,367	5,367
	R-squared	0.095	0.104	0.106

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses.

† We have also estimated the model including the time-fixed effects. The results are largely similar.

Table 5: Total Treatment Effects across Cohorts

Cohort	All months	Month 2	Months 3-4	Months 5+
Feb. 2016	26.53	35.36	31.31	25.49
	(2.02)	(2.46)	(5.52)	(2.14)
Mar. 2016	25.68	43.05	26.58	28.29
	(2.42)	(4.47)	(2.88)	(2.51)
Apr. 2016	27.45	42.00	28.56	26.43
	(1.75)	(4.31)	(2.37)	(1.75)
May. 2016	26.52	35.47	25.60	26.04
	(1.68)	(2.76)	(1.76)	(1.69)
Jun. 2016	33.52	37.89	35.05	32.84
	(2.00)	(3.15)	(3.13)	(2.03)
Jul. 2016	26.82	33.03	27.74	26.39
	(1.39)	(2.34)	(1.91)	(1.36)
Aug. 2016	35.80	37.60	32.34	33.97
	(3.95)	(4.72)	(4.03)	(4.33)

Notes: Standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Figure 1: Customer Purchases of Members vs. Non-Members

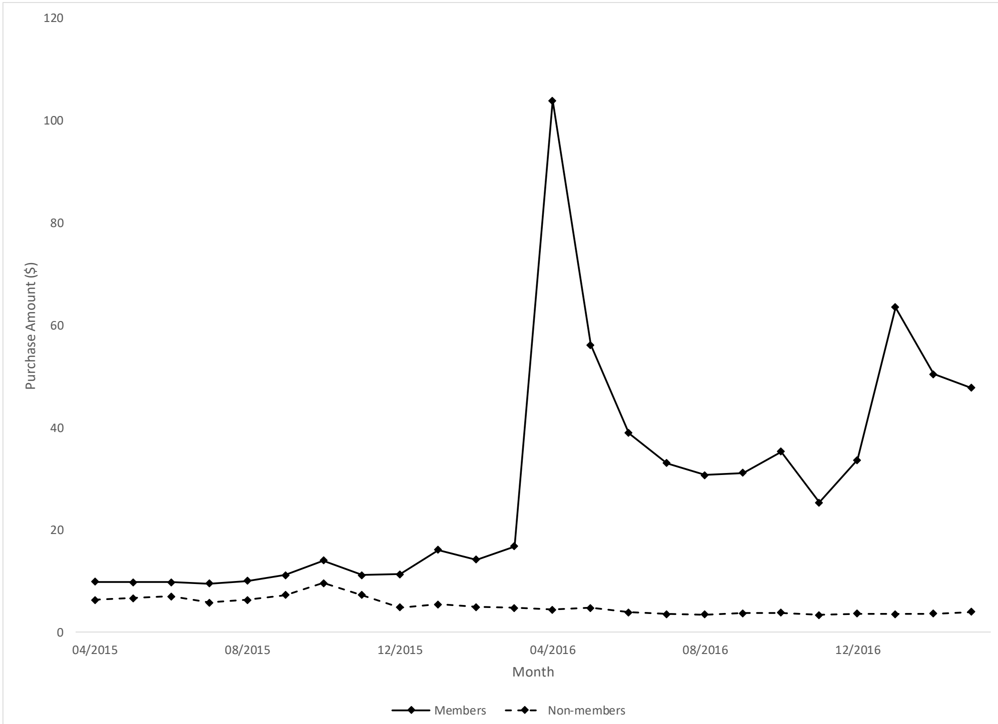
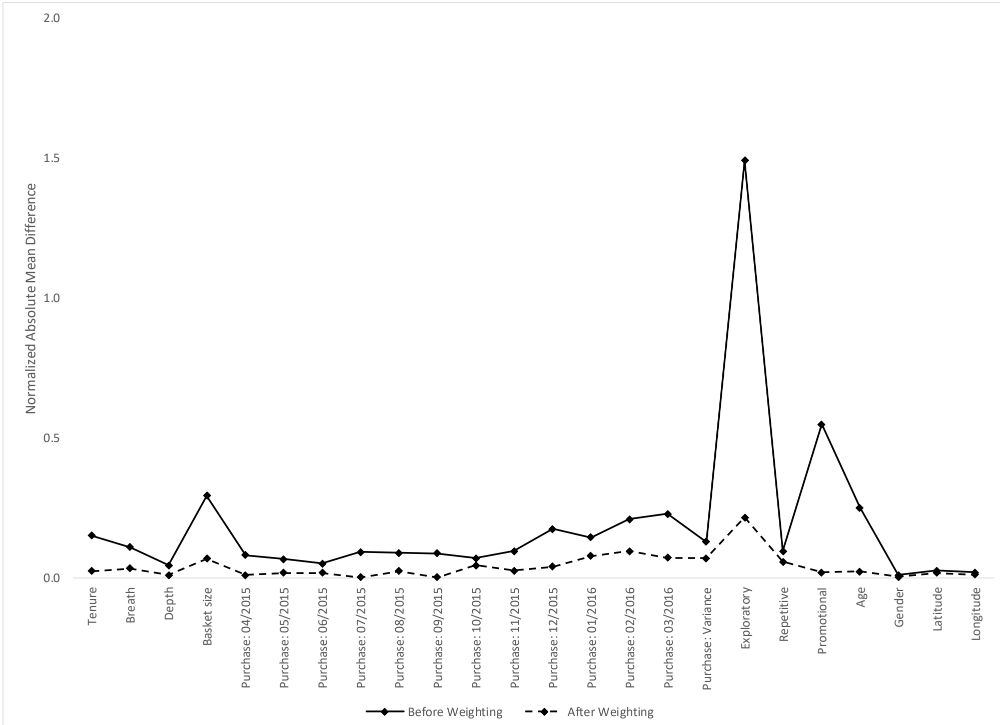


Figure 2: Covariate Balance Before and After Weighting



Notes: For the purpose of illustration, we plot the normalized absolute mean difference of monthly purchase amount along with other observed characteristics. In the matching procedure, however, monthly purchase amount from all product categories were used.

Figure 3a: Distribution of the Total Treatment Effect

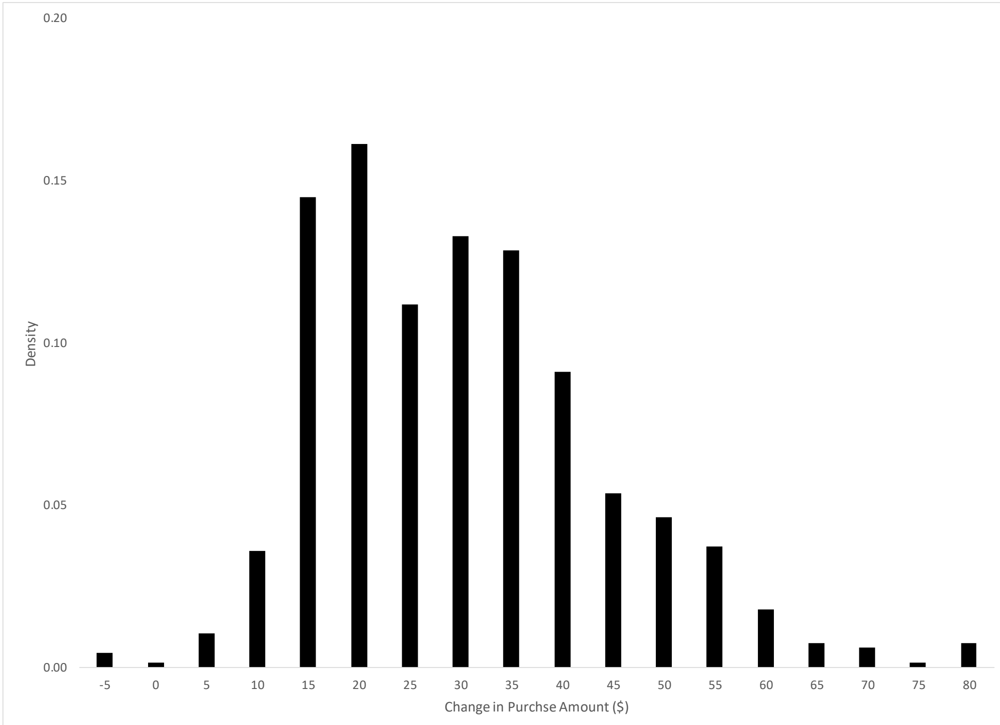
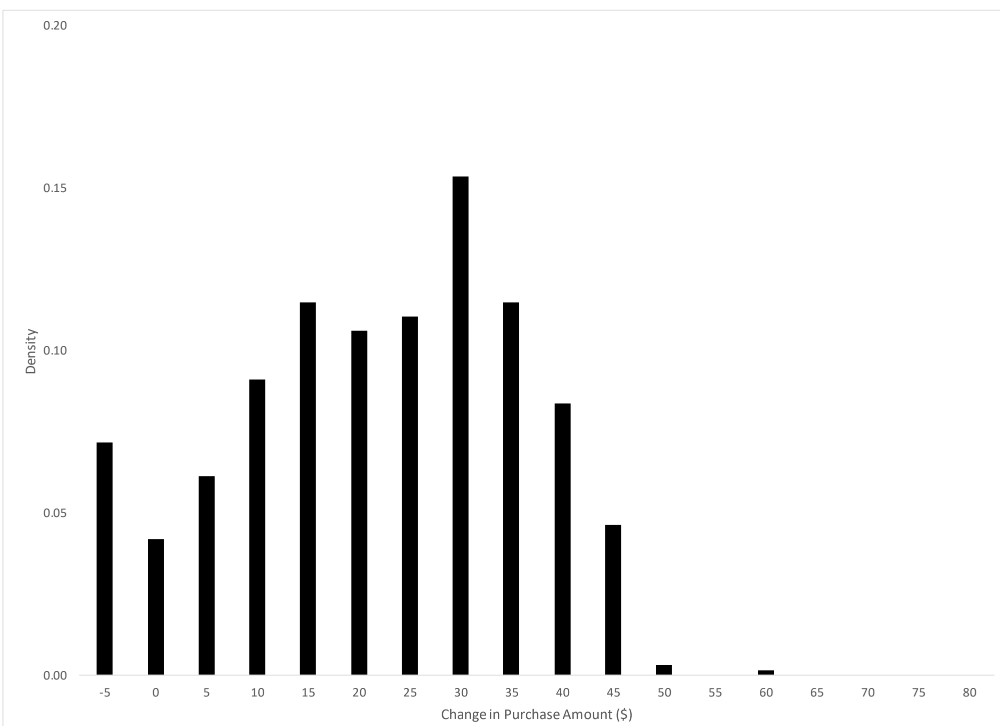


Figure 3b: Distribution of the Net Treatment Effect



Appendix A: Estimation

We describe the details for estimation, including the implementation details for GRF and the construction of the average treatment effect estimator.

A1. Implementation Details for GRF

Recall that our goal is to recover the heterogeneous parameters in Equation (2) conditional on covariates x. Let $\theta(x) = \{\tau(x), \beta_1(x), \beta_2(x)\}$ denote the parameters of interest. The GRF algorithm takes a three-step procedure. To accommodate the fixed effects and enhance the robustness of the results, we first pre-process the data. We next build a random forest that defines the weighted set of neighboring customers for each member. Finally, the treatment effects are obtained by estimating Equation (2) on the weighted sets. Below we describe each step in detail.

To accommodate the fixed effects, we first demean the variables (outcome variable, membership indicator, and marketing mix variables) in Equation (2). To enhance the robustness the results, following Athey et al. (2019), we further implement an orthogonalization procedure. That is, we residualize the (demeaned) variables by the covariates x using separate trained regression forests. The final GRF is trained on the residuals instead of the original variables. Athey et al. (2019) find orthogonalization is essential for observational studies.

To build the random forest, we first grow a decision tree through iteratively partitioning the data into subgroups (leaves). In each iteration, the algorithm chooses the covariate and the best cutoff for dividing a parent node P into two child nodes C_1 and C_2 by optimizing the Δ-criterion:

$$\Delta(\mathrm{C}_1, C_2) = \frac{n_{c_1} n_{c_2}}{n_P^2} \left| \hat{\theta}_{C_1} - \hat{\theta}_{C_2} \right|^2,$$

where $\hat{\theta}_{C_1}$, $\hat{\theta}_{C_2}$ are estimation results from Equation (2) computed in the child nodes C_1 and C_2. n_{c_1}, n_{c_2}, and n_P are number of observations in the child nodes and parent nodes, respectively. The Δ-criterion approximates the expected squared error from Equation (2) and also measures the increase in the heterogeneity of the estimated treatment effects. The algorithm continues to partition the data until there is at least k treated and controls in each leaf.

As a single tree is likely to overfit the data, an ensemble of B trees is generated. The bth tree is constructed using a random subsample without replacement containing n_b observations from a total of N observations in the data. For each tree, a random set of proportion p of the covariates can be potentially used to form the splits. The number of trees B, the minimum number of treatment and control observations in each leaf k, the subsample size n_b, and the size of the set of covariates used to build each tree p are the hyper-parameters of a forest which we choose by cross-validation. Finally, in the context of panel data, we account for clusters at the individual level in the sampling as well as estimation process.

Given the forest, we can define for each member a weighted set of its neighboring customers (which may include both non-members and other members). Specifically, for a member with covariates x, the weights are the frequency with which each customer falls into the leaves that contains x.

With the weighted set of neighboring customers defined for each member, the individual-level treatment effect is estimated by fitting Equation (2) using the weighted set of observations:

$$\theta(X_i = x) = \underset{\theta=\{\tau,\beta_1,\beta_2\}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T} w_i(x)[Y_{it} - \tau(x)M_{it} - \beta_1(x)\log(P_{it}) - \beta_2(x)\log(B_i + G_{it})]^2,$$

where the weights $w_i(x)$ measure the similarity of customer i and x. Y_{it} is the outcome variable of interest. M_{it} is an indicator for membership status that equals one if customer i is a member in month t and zero otherwise. P_{it} is the store-level price encountered by customer i in month t. B_i is the baseline spend of customer i, and G_{it} is the amount of gift card offered to customer i in month t. As noted earlier, Y_{it}, M_{it}, $\log(P_{it})$, and $\log(B_i + G_{it})$ are residualized by the fixed effects and covariates x.

We use monthly transaction data over a 24-month period. Of these, the first 12 months are prior to their subscription. We excluded customer purchases in the month of adoption to avoid any potential simultaneity bias with the adoption itself.

A2. Doubly Robust Estimator

In order to obtain the treatment effect across members and examine the average effectiveness of the program, we construct a doubly robust average treatment effect estimator.

One naïve estimator for the average treatment effect can be obtained by simply averaging individual-level effects from the GRF, i.e., for n customers with covariates X_i, $i = 1, \dots, n$, $\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}(X_i)$. The idea of the doubly robust estimator is to combine the individual-level estimates from the GRF and a regression-based prediction such that it is robust to misspecification of either the matching model or the regression model (Chernozhukov et al. 2018). Specifically, the doubly robust estimator augments the naïve estimator with the residuals from a regression forest as a bias-correction term and is defined as:

$$\bar{\tau}^{DR} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}(X_i) + \sum_{i=1}^{n}\sum_{t=1}^{T}\frac{(Y_{it}-\hat{Y}_{it})(M_{it}-\widehat{M}_{it})}{(M_{it}-\widehat{M}_{it})^2},$$

where $\hat{Y}_{it}$ and $\widehat{M}_{it}$ are predictions of Y_{it} and M_{it} from regression forests. The variance of this estimator, as is standard for linear models, is defined as:

$$Var(\bar{\tau}^{DR}) = \frac{1}{(nT-1)nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\frac{(Y_{it}-\hat{Y}_{it})(M_{it}-\widehat{M}_{it})}{(M_{it}-\widehat{M}_{it})^2}\right]^2.$$

Web Appendix A: Store-level Price

In the main analysis, we construct the store-level price ($Price_{it}$) for customer i at time t as the weighted average price of a basket of products, which is defined as follows:

$$Price_{it} = \frac{\sum_{j \in J_t} P_{gjt} \cdot Q_{jt}}{\sum_{j \in J_t} Q_{jt}},$$

where P_{gjt} is the minimum price of product j paid by group g (members or non-members) at time t, Q_{jt} is the quantity of product j purchased by members, and J_t is the set of products that were purchased by members and non-members at time t.

We investigate a few alternative ways to construct the store-level price measure. First, we define the group-level price for a product as the average price paid by the group for the product. We find the results are quite similar to those reported in the manuscript.

We also consider alternative definition for the basket of products. For example, Dubé et al. (2018) calculate a store-level price index and impute shelf price as non-promotion price when sales quantity is zero and shelf price is not observed. Following the literature, we construct a basket of products that include all the products offered by the website and impute missing prices with listed prices. We find with this measure the price variation across members and non-members is masked by the noise through the imputation of price information. This is due to the fact that there is a large number of products offered at the website and price information is very sparse in our e-commerce context. Our choice of the basket of products in the main analysis, which are those products purchased by both members and non-members, results in a store-level price that captures the price variation across members and non-members well while ensuring a sufficient number of products are included for the analysis.

Finally, we consider alternative definitions for the weights. For example, we define the weights as quantity based on purchases by both members and non-members. We find that with these weights, the store price does not capture sufficient variation in prices between members and non-members. While the net treatment effects estimates are qualitatively similar when alternative weights are specified, our choice of weights in the main analysis, the quantity purchased by members, is motivated by two reasons. First, by specifying quantity purchased by members as weights, we put slightly more weights on the products preferred by the members, which balances the need to capture sufficient variation in prices between members and non-members. Second, the resulting store price measure gives an upper bound of the price discounts encountered by the members and result a conservative estimate of the net effect of the program.

Web Appendix B: Identification of the DD Model with Covariates

We use a simulation study to understand how different patterns in purchase data can help identify the parameters of the DD model in Equation (2). As described in Section 4, the price coefficient (β_1) is identified by the month-to-month variation in purchase amount and store-level price. The parameter β_2 is identified by the difference of the unexplained increase in purchase amount upon subscription across members who differ in terms of the baseline spend. The parameter τ is identified by the remaining change in member purchases upon subscription.

For the purpose of illustration, we simulate the purchase patterns of two members with different levels of baseline spend and one non-member under various sets of parameters. Specifically, we choose two levels (low versus high) of the baseline spend among members, and set $12 for the low-spend member and $20 for the high-spend member, and $6 for the non-member. We use the values of the simulation parameters which are chosen to mimic the value found in our real data. The variation in prices (between the two groups and over time) mimics our data while the levels of prices are normalized by the average price for the non-members. All the time-fixed effects are assumed to be zero and customer-fixed effects are set to be each customer's baseline spend. Further simulations not reported suggest that parameter recoverability is robust.

Figures WB1 to WB4 show the purchase patterns over a 24-month period, excluding the month of adoption (April 2016), under different sets of parameters. Figure WB1 shows the purchase patterns when all three parameters are set to 0, i.e., $\tau = \beta_1 = \beta_2 = 0$. When customers are insensitive to economic benefits (e.g., price, gift card) and membership of the program, as expected, purchase patterns are constant at the level of baseline spend and do not change over the data period. Figure WB2 shows the purchase patterns when $\tau = 0$, $\beta_1 = -3$, and $\beta_2 = 0$. When customers are sensitive to price but do not respond to other types of economic benefits (e.g., gift card) and membership, we find there is a month-to-month variation in customer purchases that are correlated with the changes in price due to discounts. The purchase patterns also vary by customers because members obtain, for example, member-exclusive discounts post subscription, which make members purchase more than non-members who could not obtain such offers. Figure WB3 shows the purchase patterns when $\tau = 0$, $\beta_1 = 0$, and $\beta_2 = 3$. This is the case when customers are responsive to gift card but do not respond to price and membership. With parameter $\beta_2 > 0$, members increase their purchases upon subscription. Importantly, the magnitude of the increase differs depending on the member's baseline spend. Finally, Figure WB4 plots the case when $\tau = 18$, $\beta_1 = 0$, $\beta_2 = 0$. A positive membership effect results in an increase in purchase among members at the time of their subscription. And the magnitude of the increase is common across all members.

In summary, parameter recoverability is robust. We exploit the panel structure of our data and heterogeneity in the baseline spend. The parameters β_1 and β_2 are identified by the month-to-month variation in customer purchases and heterogeneity in the pre-subscription baseline spend among members, respectively. The parameter τ is identified by the remaining change in member purchases upon subscription.

Figure WB1: Customer Purchases When $\tau = \beta_1 = \beta_2 = 0$

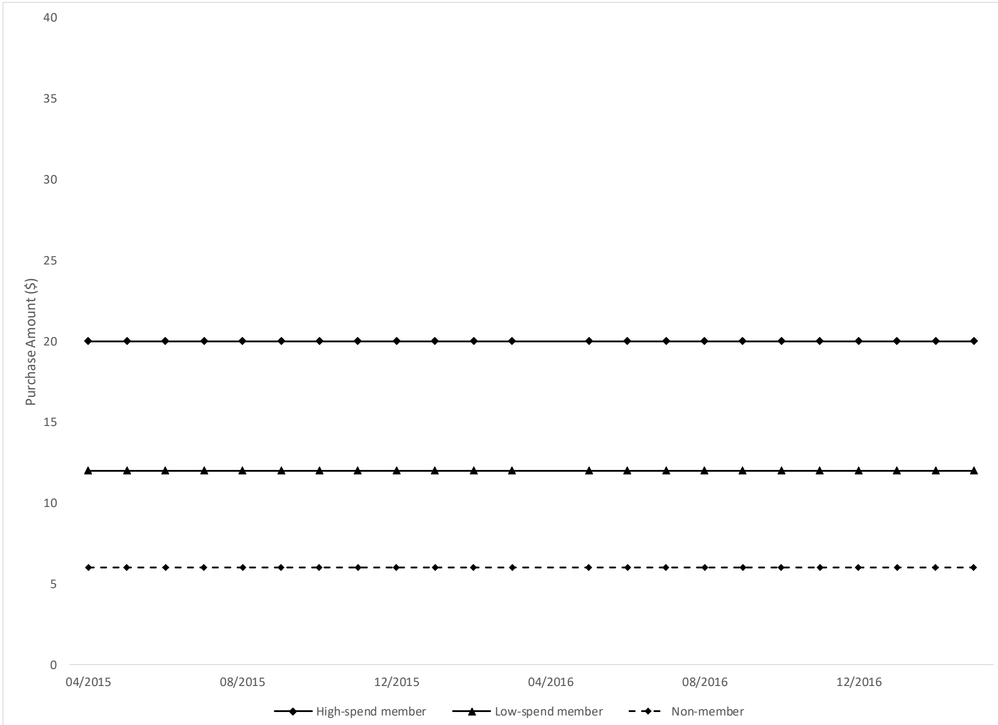


Figure WB2: Customer Purchases When $\tau = 0$, $\beta_1 = -3$, $\beta_2 = 0$

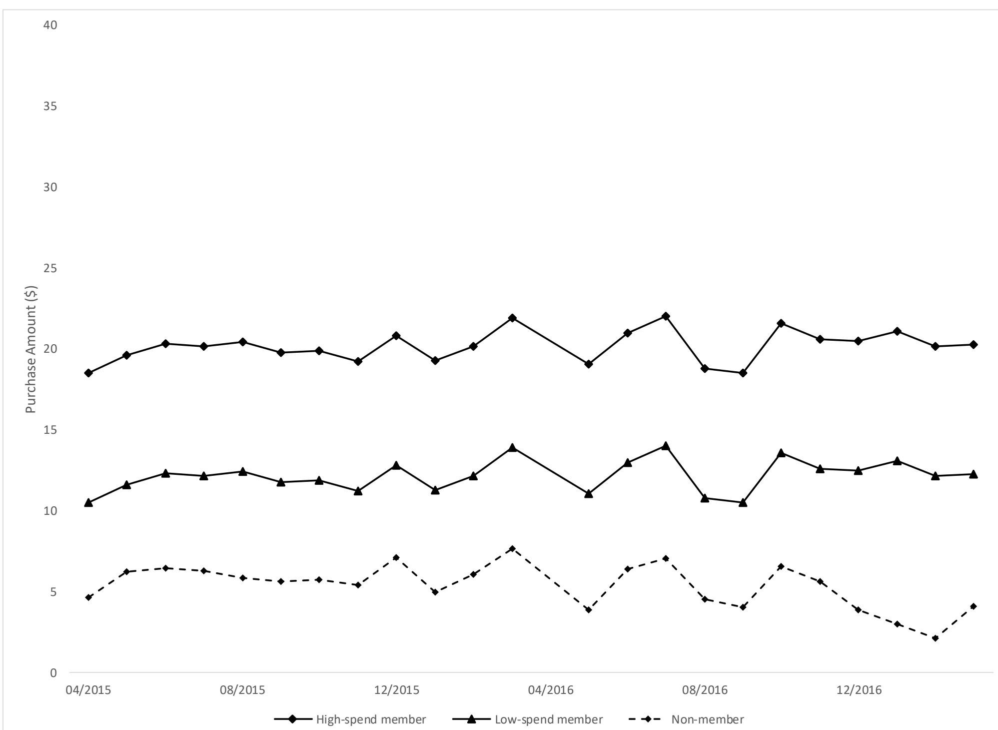


Figure WB3: Customer Purchases When $\tau = 0$, $\beta_1 = 0$, $\beta_2 = 3$

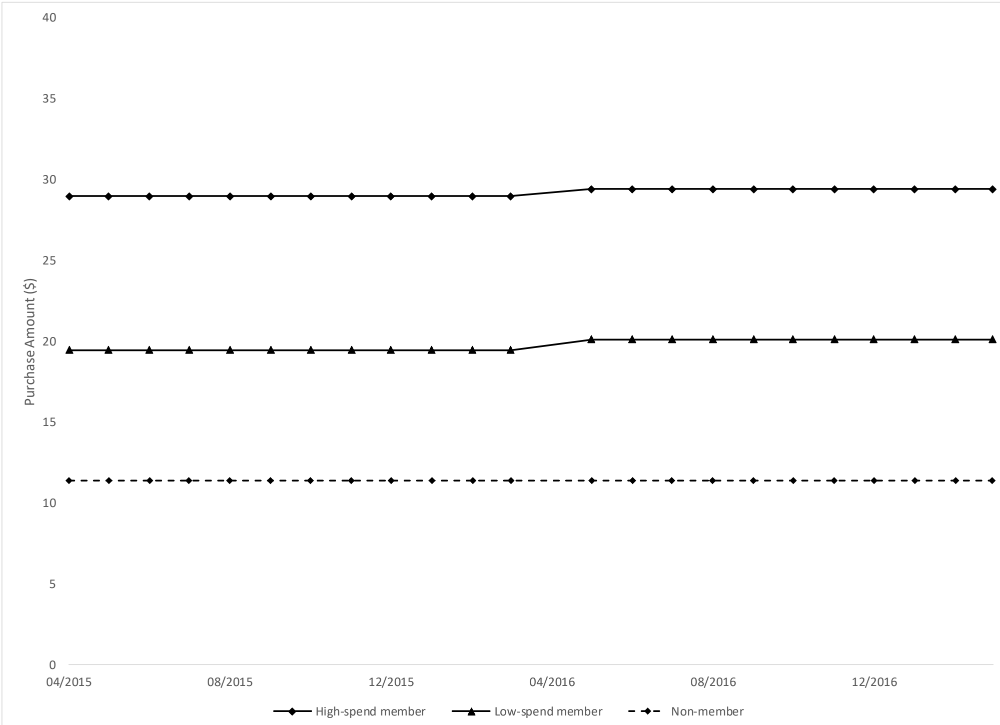
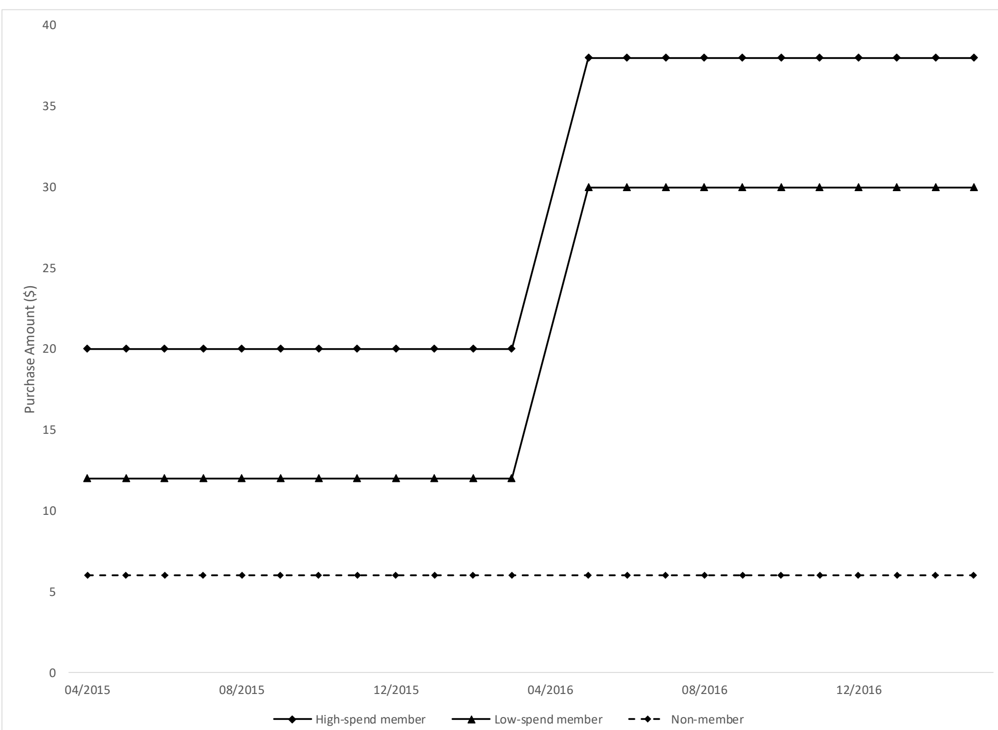


Figure WB4: Customer Purchases When $\tau = 18$, $\beta_1 = 0$, $\beta_2 = 0$



Web Appendix C: Placebo Tests

Table WC1 reports the results from the placebo tests. We find that along all dimensions of interest, members' purchase behavior does not change six months prior to subscription. These results indicate that the members and the matched non-members have statistically comparable pre-treatment time trend.

Table WC1: Placebo Test

	Placebo Effect
Purchase amount ($)	1.50
	(1.49)
Purchase frequency	0.03
	(0.03)
Basket size	1.20
	(2.93)

Notes: Robust standard errors appear in parentheses.

Web Appendix D: Identification of Heterogeneous Treatment Effects

The identification of heterogeneous treatment effects relies on the same source of variation in the data as those used to identify the DD model discussed in Web Appendix B. However, one major challenge in estimating heterogeneous treatment effects is that the individual-level treatment effects are estimated on a weighted (matched) sample which is potentially of a smaller size than the full sample. If there were not enough variation in these weighted samples, the individual-level treatment effects would be estimated imprecisely. Hence, we use a simulation study to show that there is sufficient variation in our data to identify the heterogeneous treatment effects.

To demonstrate the recoverability of the parameters, we simulate a data to mimic our data set in size and nature (e.g., number of members and non-members, length of the data for each customer, marketing mix, covariates). We also allow the model parameters to be non-linear functions of the covariates. The fixed effects are set to be zero. Finally, the error terms are drawn from the standard normal distribution to generate the outcomes.

We apply our proposed estimation method to estimate heterogeneous treatment effects. Figures WD1 to WD3 compare the estimates with the true parameters across all individuals. We find the recovery of the individual-level estimates is robust and suggests that there is sufficient variation in the data to identify heterogeneous treatment effects.

Figure WD1: Parameter Recovery of τ

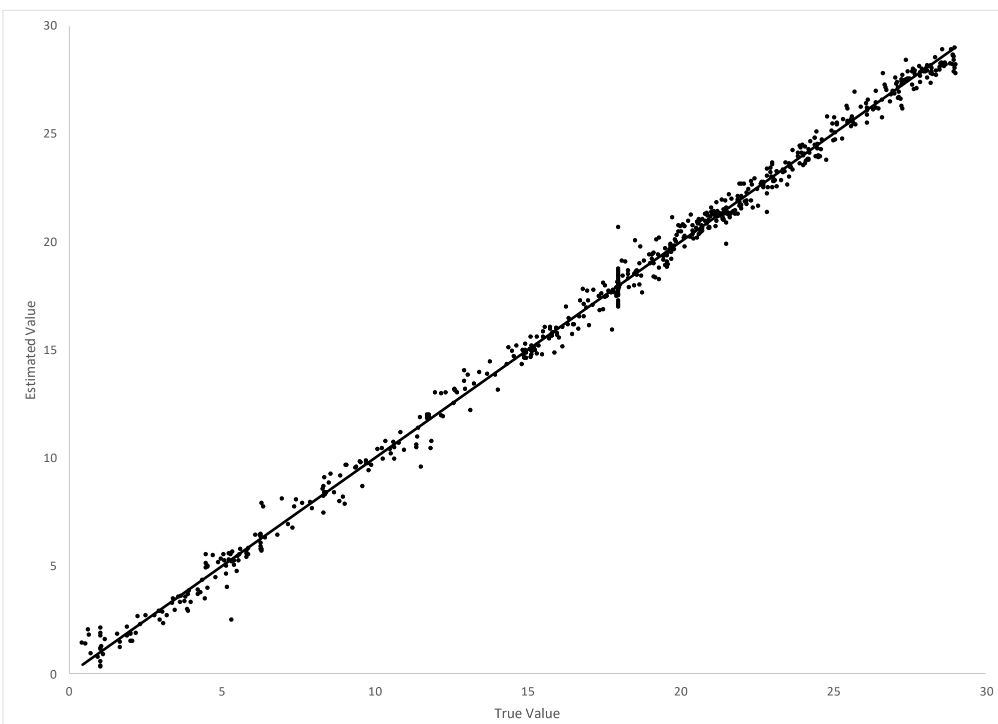


Figure WD2: Parameter Recovery of β_1

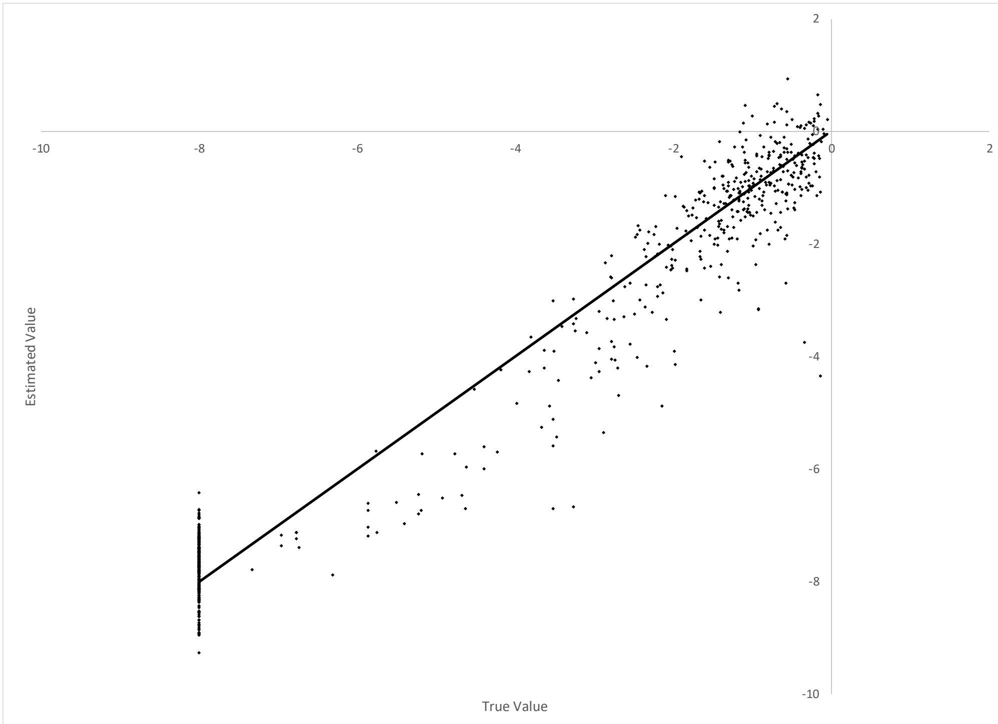
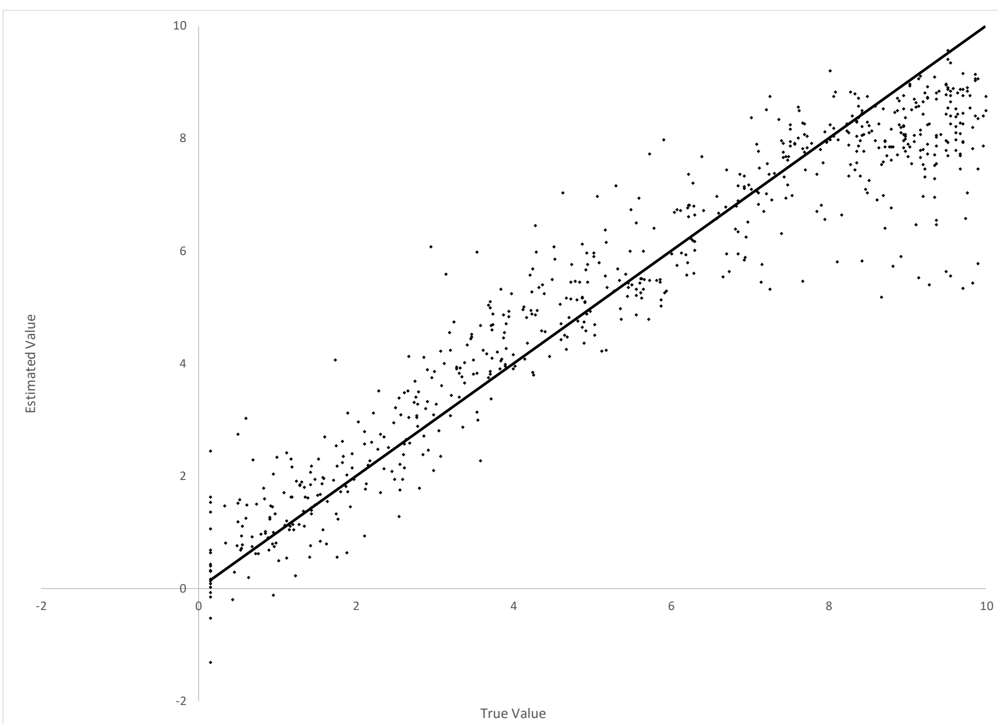


Figure WD3: Parameter Recovery of β_2



Web Appendix E: Average Treatment Effects with $\log(Y_{it})$

Tables WE1 and WE2 report the average total and net treatment effect on customer purchases, respectively, where the outcome measure was transformed with natural log in the estimation.

Table WE1: Total Treatment Effect with $\log(Y_{it})$

	All months	Month 2	Months 3-4	Months 5+
Purchase amount ($)	1.57	2.05	1.80	1.50
	(0.05)	(0.13)	(0.08)	(0.27)
Purchase frequency	1.47	1.87	1.57	1.67
	(0.11)	(0.13)	(0.12)	(0.15)
Basket size	-0.34	-0.17	-0.24	-0.33
	(0.14)	(0.09)	(0.05)	(0.11)
Variety ($)				
Purchase amount in known categories	1.82	2.08	1.86	1.63
	(0.05)	(0.11)	(0.08)	(0.06)
Purchase amount in new categories	0.19	0.15	0.24	0.18
	(0.02)	(0.03)	(0.03)	(0.02)
Purchase amount of known products	0.08	0.06	0.05	0.05
	(0.04)	(0.04)	(0.04)	(0.04)
Purchase amount of new products	1.92	2.26	2.13	1.77
	(0.04)	(0.09)	(0.06)	(0.04)

Notes: Standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table WE2: Net Treatment Effect with $\log(Y_{it})$

	All months	Month 2	Months 3-4	Months 5+
Purchase amount ($)	1.34	1.63	1.49	1.11
	(0.05)	(0.11)	(0.08)	(0.05)
Purchase frequency	0.47	0.54	0.54	0.40
	(0.01)	(0.04)	(0.02)	(0.01)
Basket size	-0.24	-0.16	-0.24	-0.29
	(0.17)	(0.09)	(0.04)	(0.12)
Variety ($)				
Purchase amount in known categories	1.32	1.58	1.46	1.13
	(0.05)	(0.11)	(0.07)	(0.06)
Purchase amount in new categories	0.15	0.12	0.21	0.17
	(0.03)	(0.05)	(0.02)	(0.04)
Purchase amount of known products	0.07	0.11	0.04	0.05
	(0.03)	(0.04)	(0.03)	(0.05)
Purchase amount of new products	1.42	1.64	1.51	1.27
	(0.04)	(0.08)	(0.05)	(0.05)

Notes: Standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.