The Impact of Subscription Programs on Customer Purchases

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

September 2021

* 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 Assistant Professor of Marketing at Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore; email: qiyu@smu.edu.sg. 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 in 2017, 2018 and 2020 ISMS 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 have become increasingly popular among a wide variety of retailers and marketplace platforms. Subscription programs give members access to a set of exclusive benefits for a fixed fee upfront. In this paper, we examine the causal effect of a subscription program on customer behavior. To account for self-selection and identify the individual-level treatment effects, we combine a difference-in-differences approach with a generalized random forests procedure that matches each member of the subscription program with comparable non-members. We find subscription leads to a large increase in customer purchases. The effect of subscription 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 two thirds is attributed to the non-economic effect. We provide evidence that members experience a sunk cost fallacy due to the upfront payment that subscription programs entail. Finally, we present the profitability of the subscription program and 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

In an effort to retain and develop customers, retailers and marketplace platforms 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 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 similar programs with benefits that 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] However, 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 change their purchase behavior due to the economic benefits they receive post subscription. Third, mere membership can also bring value and change customer behavior, for instance, by leading them to form a new consumption habit or to feel enhanced status. The industry report cited above indicates that 82% of Prime

[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 program 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 in retail.
[2] "Here's How Much Amazon Prime Customers Spend Per Year," available at http://fortune.com/2017/10/18/amazon-prime-customer-spending/.

members shop with Amazon even when the price is lower elsewhere, suggesting that the impact of subscription programs can go beyond the economic benefits offered. From a theoretical perspective, it is important to identify the economic and non-economic effects of subscription on customer behavior separately. While the economic effect of a program may be specific to the features of the program, the non-economic effect due to the underlying psychological drivers is likely to be applicable in other contexts. Determining the relative contribution of the two components is substantively important to improving the design of subscription 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 purchase behavior. We hereafter refer to the impact as the treatment effect of subscription. We also seek to decompose the treatment effect into the economic effect due to program benefits and the non-economic effect that cannot be explained by the tangible benefits of the program (e.g., status from the program). Specifically, we are interested in addressing the following questions: Does a subscription program generate value for a firm? Is the subscription program effective in inducing customers to change their behavior because of the economic benefits and/or the psychological drivers? How do these effects vary over time and across customers? What are the underlying drivers of any documented effects? We address these questions in close collaboration with a company that launched a subscription program at its online channel. The program offers members a few exclusive benefits for an upfront fee. Our data contains individual-level transactions before and after the launch of

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

the program and other information on various components in the program, thus allowing us to examine the effect of subscription 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 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 and time fixed effects to estimate the 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 member based on a large set of observed characteristics following a generalized random forests (GRF) procedure (Athey et al. 2019) and estimate the DD model using the weighted sample. The combination of the DD approach and GRF procedure 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 non-economic effect by evaluating the residual effect of the program after controlling for the marketing mix a member was exposed to. The individual-level estimates of economic and non-economic effects allow us to get a richer understanding of the impact of the subscription program and its underlying drivers.

We find subscription is effective in lifting sales. On average, members increased their purchases by about $27 per month over a 12-month period post subscription, which is more than double of customer monthly purchase amount prior to subscription. The treatment effect of subscription is economically significant and persistent over time. The subscription 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 due to the economic benefits of the program and the remaining two thirds is attributed to the non-economic effect. There is also a large variation in the treatment effect across customers. Our main findings are robust to potential confounding effects of self-selection and unobservables, different treated groups, and different outcomes of purchase behavior.

We investigate the potential drivers at work that explain our findings. We find that, in addition to the psychological underpinnings (e.g., affect, habit, status) documented in the context of other types of membership programs, a unique feature of subscription programs helps lift sales: 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. Further, we provide the profitability of the subscription program.

Our paper is related to several streams of research. First, we contribute to the literature on subscription business. 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 stand-alone 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 program benefits to justify their subscription decisions. As the upfront fee is a feature common to subscription programs, we believe our findings 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, Gopalakrishnan et al. 2021). Others find no or weak evidence loyalty programs are effective (e.g., Hartmann and Viard 2008). Several researchers have documented loyalty programs can induce positive affect (e.g., Leenheer et al. 2007), lead to the development of habitual consumption (e.g., Wood and Neal 2009), and enhance members' perception of status (e.g., Drèze and Nunes 2009). This paper contributes to this literature by using quasi-experimental data to measure the effect of subscription 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 its extensive evidence 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. Sections 6 presents several robustness checks. Section 7 discusses the profitability of the subscription program. We conclude in Section 8.

2. Subscription Business

A subscription-based business is one in which a customer periodically 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, an adoption 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]

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. Customers benefit from this type of subscription

[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.

[5] 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.

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. Examples include Amazon (Prime), Barnes & Noble (B&N Membership), Sephora (Flash), Alibaba (88VIP), etc. Access subscriptions differ from the other two described above in that a subscription program is usually initiated by an existing non-contractual business and provides members exclusive benefits beyond those available to regular customers (i.e., non-members).[6] Thus, access subscriptions have a very wide appeal, as they can be adopted by nearly all business-to-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.

[6] Access subscriptions can also be a stand-alone business in service subscriptions.

While there is anecdotal evidence suggesting the commercial success of subscriptions, no study has evaluated yet 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 economic benefits associated with the program and/or the non-economic benefits. For the latter, 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.[7] Our research aims to fill in this gap.

3. Empirical Context and Data

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). It has significant brick-and-mortar and online presence, with the latter being smaller than the former. In December 2015, a subscription program was launched on its e-commerce website. The launch of the program and its benefits were communicated through mass emails and on their website, and no targeting was involved. The program provided both unlimited use of a service and access to member-exclusive offers for a subscription fee of $50 per year. Specifically, 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. Several products were occasionally coupled with member-exclusive discounts. Free samples were offered to members monthly with a purchase online. Finally, members also had access to unlimited free shipping with no minimum purchase requirements. Similar to other subscription

[7] "The In Crowd: Inside Boston's Elite Country Clubs," available at https://www.bostonmagazine.com/news/2018/09/11/in-crowd-country-clubs/.

programs, these benefits were offered beyond the first year of subscription rollout. Other than the program benefits mentioned above, there is no difference in the communication between members and non-members during the data period.

Our data include 10,811 customers who joined the subscription 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 342 to 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 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 the 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 address, which we utilize to control for customer heterogeneity while explaining the drivers of purchase behavior.

Using transaction data, we define a set of outcome measures associated with customer purchases. As the program was offered at the online channel only, unless specified otherwise, they are based on online purchases and are constructed at the customer-month level, which is the unit of analysis in this research. As our main 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. These two measures, while both positively correlate with the amount spent, may have differing implications for the firm in

[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.

terms of engagement and costs. We also characterize the variety in purchase behavior with a few metrics. We classify a product (and its category) a customer purchased as a new versus known product (category) on the basis of whether 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: 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 post subscription.

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 implementation details.

4.1 Empirical Strategy

A key challenge in identifying the impact of subscription on purchase behavior is due to self-selection—members may differ from non-members even before they subscribe the program and this may lead to biased results if we estimate the effect by directly comparing purchases between members and non-members. We also seek to identify the economic and non-economic effects of subscription on customer behavior separately. The former captures the changes in purchase behavior attributable to the tangible benefits of the program (e.g., reduced prices due to member-exclusive discounts) and the latter includes any remaining effect on demand. Finally, we are interested in examining the heterogeneity in the treatment effect across customers.

[9] Based on conversations with the retail partner, we decided to have five product categories for our empirical analysis 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, brushes, and accessories).

To control for self-selection and identify the 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 time-invariant unobservables. Within the regression framework, we separately identify the non-economic effect on purchase behavior by evaluating the residual effect of the program after controlling for the marketing mix a member was exposed to. Conceptually our framework is similar in spirit to past work that has examined how the different components of a pricing scheme may have an impact on demand over and above its economic effects (e.g., Bertini and Wathieu 2008, Iyengar et al. 2011). We complete our modeling framework by embedding the DD specification within a generalized random forests (GRF) 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-dimensional 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 effect of the subscription program, as it gives well-defined pre-treatment 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 Section 6, our findings are robust to different treated groups who joined the program during other months.

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

Before we establish the 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 (April 2015 to March 2016) are prior to their 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 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. = 3.09, $p < 0.001$). Clearly, it will be biased if we estimate the effect of subscription by comparing customer purchases between the two groups.

Insert Table 1 about here

4.3 Difference-in-Differences

The descriptive analysis reveals that members purchased considerably more than non-members post subscription. However, this analysis may suffer from self-selection. In this section, we employ a DD approach which controls for time-invariant unobserved variables and quantify the (causal) treatment effect of the subscription program. We estimate the following DD model:

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

where Y_{it} is the outcome measure of customer i in month t and $Member_{it}$ 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 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 parameter of interest is τ, which captures the treatment effect (sum of economic and non-economic effects) on purchase behavior.

To measure the non-economic effect of the program, we extend Equation (1) and further control for the marketing mix that a member was exposed to. In our context, members received unlimited free shipping service and member-exclusive offers. We find free shipping and product samples had limited impact.[11] Members benefited from price discounts and gift cards: members had an average 6% additional discount at the store level as compared to non-members and about 20% of monthly gift cards were redeemed.

To model the economic effect of the program, consider a customer who determined her purchase behavior given the store-level price and her monthly budget, which may be altered by monthly gift cards. We specify a log-linear model in which the customer's purchase is linear in the logarithm of the price and the logarithm of her monthly budget. The log-linear model has been used widely in the marketing literature to characterize consumption patterns (e.g., Dubé et al. 2018), because it is consistent with utility maximization (e.g., Sato 1972).[12] Specifically,

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

[11] The firm offered customers, regardless of subscription, free shipping on orders above a certain threshold, which was satisfied by most orders. Section 5.2 provides a detailed discussion on the effect of free shipping. We also examined whether a customer purchased a product after receiving a free sample of that product, and find that less than 1% of purchases were induced by free samples.

[12] Web Appendix A discusses conditions under which our demand model is consistent with utility maximization.

where $Price_{it}$ is the store-level price for customer i in month t, $Baseline_i$ is the baseline budget of customer i, and $Giftcard_{it}$ is the amount of gift card customer i received in month t. The parameters β_1 and β_2 are the semi-elasticities of price and gift card, respectively, and the parameter τ captures the non-economic effect.

We compute the store-level price for customer i in month t ($Price_{it}$) as the weighted average price for a basket of products (e.g., Dubé et al. 2018). Given that we do not observe the prices of all products that a specific customer was exposed to, we proxy the customer-level prices using group-level (members and non-members) prices. This assumption is reasonable as there was no other source of price discrimination between members and non-members beyond the discounts from subscription. Specifically, we operationalize the monthly store-level price as the weighted average price of a basket consisting of all products purchased by both members and non-members in that month.[13] Additionally, we operationalize the baseline budget of a customer (without gift cards) by using her maximum monthly spend one year prior to subscription.

The identification of Equation (2) exploits the panel structure of our data and heterogeneity in the baseline budget. Our panel consists of data over a 24-month period, where the first 12 months are prior to subscription. 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 store-level price to identify the price coefficient (β_1). Note that all members received a gift card worth three dollars each month. The parameter β_2 is identified by the difference of the unexplained change in purchases upon subscription across members who had different baseline budgets. For example, a positive β_2 indicates that the marginal effect of the gift card is lower for members with larger budgets. The parameter τ is identified by the

[13] Web Appendix B describes how store-level price is operationalized for the main analysis and discusses the robustness of our results under alternative operationalizations.

remaining change in member purchases upon subscription and provides an estimate for the non-economic effect of the program.[14]

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 an average member increased over time even before her subscription. Thus, the parallel time trend assumption may not hold and the validity of the DD estimator is questionable (e.g., Bertrand et al. 2004).[15] 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 issues in a principled manner. We employ the generalized random forests (GRF) method (Athey et al. 2019). Similar to other methods for causal inference using observational data, e.g., kernel matching (e.g., Hastie et al. 2009), propensity score matching (e.g., Hirano et al. 2003) and synthetic control (e.g., Abadie et al. 2010), the key idea of GRF is to define for each member a weighted set of neighbors that shares similar covariates and fit the model of interest using these

[14] Web Appendix C describes how different patterns in purchase data could help identify model parameters.

[15] 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: $Y_{it} = \sum_m \tau_m \cdot Member_{i,t-m} + \alpha_i + \gamma_t + \epsilon_{it}$, where the indicator variables $Member_{i,t-m}$ equal 1 if customer i joined the subscription program in month t. 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 with marketing mix and find the parallel trend assumption does not hold. Results are available from the authors upon request.

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 for matching, 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 period (April 2015 to March 2017) because we are interested in examining the effect of subscription 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).[16]

To control for potential confoundedness, we include three sets of covariates that describe members and non-members in the pre-treatment period. The first set of covariates relates to the

[16] The estimate we report provides a conservative estimate of the effect of subscription on customer purchases. Detailed results, which include the month of subscription adoption, are available from the authors upon request.

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 categories. We also include the standard deviation of monthly purchase amount because it could relate to customer response for unlimited free shipping service.

The second set of covariates relates to customers' purchase characteristics (e.g., Baumgartner 2002). We create a few summary statistics to capture different aspects of purchase behavior: exploratory (purchases of new products), repetitive (repeat purchases of a product), and promotional (price discount rate received for purchases).[17]

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 mailing address because it can help control for other unobserved socio-demographics that affect subscription, e.g., education, income, lifestyle, 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.[18]

Insert Table 2 about here

---

[17] We conducted our analysis with alternative operationalization of the variables, e.g., share of repeat product purchases per month and category. We also used alternative thresholds to construct the variables. We find our results are largely similar.

[18] Before we built 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 join 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 had propensity scores with no counterparts in the control group and excluded them for our final estimation. We also excluded 2,023 non-members because they were not on the common support. Our findings are based on a sample of 671 members and 11,745 non-members.

We briefly describe the procedure for building the forest.[19] 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 a 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 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 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.[20]

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.[21] The procedure then fits Equation (2) with the weighted set of observations for each

[19] Web Appendix D provides the implementation details.
[20] In Appendix A, we tested whether parallel time trends hold in our context using a series of placebo tests. We find members and matched non-members have statistically comparable pre-treatment time trends.
[21] We used out-of-bag predictions to avoid over-fitting, i.e., only trees that do not include the member during tree building were used to produce the weights.

member. Hence, we effectively combine GRF with DD specification after controlling for the economic benefits of 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).[22]

5. Findings

In this section, we discuss the main findings on the treatment and non-economic effects of subscription on customer purchases. We also discuss how these effects vary over time and across customers. Finally, we explore some possible explanations underlying the effect.

5.1 Average Treatment Effects

Table 3a reports the average treatment effect of subscription on customer purchases.[23] Because our objective is to identify the effect of subscription for the long term, we first discuss the estimates over a 12-month period post subscription. The first column in Table 3a shows that on average, purchase amount per month among members increased by $27.45. The effect is economically significant, as purchase amount per month was about $12 prior to subscription.

Insert Table 3a about here

The treatment effect on purchase amount is quite striking. As shown in Figure 1, however, data patterns suggest that our finding is not an artifact. We conducted a literature review and observe the impact of membership programs has substantial variation, ranging from

[22] Web Appendix E describes how the individual-level estimates are identified.
[23] Appendix B reports the results in which outcome measures are transformed with natural log in the estimation. The results are qualitatively similar.

no effect to as high as 150% increase in customer response across various empirical settings.[24] For instance, using an 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 revenue. Using data from a men's hair-salon chain, Gopalakrishnan et al. (2021) find the introduction of a non-tiered loyalty program increases customer value by 19% over a five-year horizon. 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. 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), Kopalle et al. (2012), and Gopalakrishnan et al. (2021) study customer response to reward programs in a single category (e.g., hotel), Liu (2007) examines the impact of a loyalty program on customer behavior in a firm offering multiple categories. Our empirical context is close to Liu's because the focal firm offers a wide assortment of brands and products across different categories. The effect size in our study is considerably larger than those documented in the literature on (free) loyalty programs. Our finding 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 during their subscription. In contrast, existing literature examines customer response in non-contractual reward programs.

[24] 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.

The increase in customer purchases could be driven by the increase in purchase frequency and/or basket size. Our results show members made about one additional purchase per month (1.15) post subscription. Interestingly, we find a small but significant decrease in basket size ($-5.14). One possibility is that members might make their basket into smaller ones to utilize, for example, recurring monthly gift cards. Our results on the economic effects of 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.[25] At the category level, more than 95% of the increase in purchase amount ($26.19 out of $27.45) came from known categories. 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.

We also 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 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

[25] 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 minimizes the mean squared error, matches depend on both covariates and outcome variable. Therefore, matches can differ slightly when evaluating the effect on different outcome variables. However, unconfoundedness guarantees that all the estimates are unbiased.

subscription on purchase behavior, we utilize the purchase patterns shown in Figure 1 and distinguish between the 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 of subscription on customer purchases is the largest (an increase of $40.57) within the first two months and persisted (an increase of $25.53 per month) after four months upon subscription.

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

5.2 Non-economic Effects

We next discuss the non-economic effect of subscription for which the impact of the tangible benefits of the program is controlled. Table 3b reports the average non-economic effect of the program. Interestingly, we find about two thirds of the treatment effect on purchase amount ($17.91 out of $27.45) is due to the non-economic effect. By calculating the relative contribution of marketing mix (price discounts and gift cards), we find price discounts accounted for about $4.20 increase in purchases and gift cards explained about $5.30 increase in purchases.[26]

Insert Table 3b about here

Looking at the temporal patterns of the non-economic effect, we find the increase in purchase amount is largest (an increase of $37.28) within the first two months post subscription.

[26] We checked the face validity of the results by comparing our price and coupon elasticities to those documented in past research. We find the price elasticity (percentage change in purchase amount in response to a one percent change in price) is -1.75 and the marginal effect of coupon (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 sales.

While it gradually faded away over time, the effect is persistent and managerially important. Across most of the other metrics as well, interestingly, a significant part of the increase in customer purchases is not accounted for by the economic benefits of the program. The non-economic effect on basket size is not significant, indicating that after accounting for the economic benefits, there is no further impact on the basket size. In summary, even after controlling for the economic benefits of the program, a significant part of the effect on customer purchases can be attributed to the non-economic effect of the subscription program.

Before discussing the heterogeneity in the treatment effect, we investigate whether the increase in purchase can be attributed to the free shipping benefit associated with the subscription program. Past research has found that free shipping can be effective in increasing the order incidence and purchase amount, with the effect size varying by the savings in the shipping fee (e.g., Lewis 2006, Lewis et al. 2006). In our context, the firm offered to all its customers (regardless of subscription) free shipping on orders above a certain threshold which was satisfied by a large majority of orders. Hence, we expect that the free shipping benefit associated with the subscription program likely did not play an important role for the sales lift. Nonetheless, we present a few pieces of evidence corroborating our conjecture. First, while average basket size by members decreased a little bit, this change becomes insignificant after we account for other economic benefits of the program.[27] Second, we estimate the lower bound of the non-economic effect that excludes the effect of free shipping.[28] We find that after excluding

[27] We also replicated this finding when we examined the changes in the number of orders below the free shipping threshold. We find members had significantly more orders below the free shipping threshold post subscription. However, after taking price discounts and gift cards into account, the effect becomes insignificant.
[28] We considered the two ways in which free shipping may affect purchase amount. First, members would likely place more orders below the free shipping threshold, which would have been abandoned prior to subscription. Second, members would no longer take the effort to increase their basket sizes to meet the minimum threshold. Taken together, the effect of free shipping on purchase amount should be bounded above by the increase in purchase from the orders below the free shipping threshold. We can therefore obtain a more conservative estimate of the non-

the below-threshold orders, purchase amount by members increased significantly by $16.29 after controlling for other economic benefits. This estimate is close to $17.91, which is the magnitude of the non-economic effect reported in Section 5.2. In sum, the free shipping benefit offered through the subscription program is not likely the driver of the observed effects in our context.

5.3 Heterogeneous Treatment Effects

Figures 3a and 3b show the distribution of the treatment and non-economic effects of the subscription program 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 treatment 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 non-economic 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 treatment effect explained by the non-economic effect also has substantial heterogeneity across members. On average, 65% of the increase in purchase amount was driven by the non-economic benefit of the program. Among 15% of members, almost 90% of the increase in purchases 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 than 80% of the increases in purchase was due to the non-economic effect of the program. These

economic effect by excluding all purchases from orders below the threshold and revisit the analysis described in Section 4.

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).

5.4 Possible Explanations

In this section, we discuss the likely underlying mechanism(s) for the increase in purchase. We discuss four psychological drivers motivated by the institutional details of subscription programs and the broader literature on membership programs. In what follows, we first examine each mechanism in isolation. We summarize by confirming these insights in a unified framework.

5.4.1 Sunk Cost Fallacy

In typical subscription programs, members pay an upfront subscription fee. This unique feature distinguishes subscription programs from other types of (free) membership programs (e.g., loyalty programs). While rational customers should not take the upfront fee into account when making subsequent purchase decisions, extant work suggests people exhibit sunk cost fallacy, tendency to hold the initial (sunk) cost in a mental account and change their behavior to amortize the psychological burden of the cost (e.g., Thaler 1980). In the context of subscription programs, we posit that the upfront fee of the program creates a sunk cost and induces an increase in customer purchases such that they can utilize the program benefits (e.g., price discounts) and recover the initial payment. Following this logic, we hypothesize that as customers benefit more from the program, sunk cost is amortized and its effect wears off.

To test this hypothesis, we first obtain the estimates of the non-economic effect at the member-month level ($\hat{\tau}_{it}$). We extend the DD model in Section 4 and augment it with GRF:

$$Y_{it} = \sum_{s=1}^{11} \tau_s Member_{i,t-s} + \beta_1 \, log(Price_{it}) + \beta_2 \, log(Baseline_i + Giftcard_{it}) + \alpha_i + \gamma_t + \epsilon_{it}, \quad (3)$$

where $Member_{i,t-s}$ equals one if customer i joined the subscription program in month $t-s$ and zero otherwise, i.e., $Member_{i,t-s}$ is the variable which indicates s months since subscription. In essence, $\hat{\tau}_{is} = \tau_s(X_i)$ estimates the change in purchase in month s relative to the pre-subscription period and captures the non-economic effect in month s. To test whether the non-economic effect is moderated by the usage of program benefits, we estimate the following model:

$$\hat{\tau}_{it} = \gamma \cdot X_{it} + \delta_i + \xi_{it}, \tag{4}$$

where X_{it} is the cumulative program benefits utilized by member i prior to month t. We also include individual-level fixed effects to account for time-invariant heterogeneity across members. As the dependent variable is estimated with error, we report heteroskedastic-robust standard errors clustered at the individual level to account for the variance of the dependent variables (e.g., Hanushek 1974, Liang and Zeger 1986). The first column in Table 4 reports the result from the regression. The results indicate that consistent with sunk cost fallacy, as customers utilized more program benefits and recovered their sunk cost, the non-economic effect of the program declined.[29]

Insert Table 4 about here

By construction, the dependent and independent variables in Equation (4) may display systematic time trends; the non-economic effect may likely decline while the cumulative benefits monotonically increased over time. These time trends may result in a spurious correlation between the non-economic effect and the cumulative benefits. We detrend the dependent and independent variables through first differencing (e.g., Hamilton 2020) and estimate the following regression:

$$\hat{\tau}_{it} - \hat{\tau}_{i,t-1} = \gamma \cdot (X_{it} - X_{i,t-1}) + e_{it}. \tag{5}$$

[29] We also investigated the presence of sunk cost fallacy by examining whether members decelerated their purchases after using \$50 gift cards received upon subscription. We do not find a significant decrease in customer purchases. As most customers used their \$50 gift cards in the very first month upon subscription, there was limited variation to detect an effect.

In essence, the first difference model tests whether more benefits used in month $t-1$ lead to larger decline in the non-economic effect in month t. As shown in the second column in Table 4, we find that consistent with sunk cost fallacy, the increase in purchase is smaller when customers utilized more program benefits.

5.4.2 Affect

Past literature on membership programs suggests that a membership program can induce a positive affect towards the firm and lead to increased purchases (e.g., Leenheer et al. 2007). Our prediction based on positive affect is that customers act favorably towards the program shortly after they become a member but experience hedonic decline as they continue to purchase from the firm (e.g., Galak and Redden 2018). Thus, we would expect the impact of the program to decrease with past purchase amount. To test this prediction, we estimate Equations (4) and (5) using the cumulative purchase amount as the explanatory variable. Similar to our analyses presented above, we estimate the fixed effect model and account for the common time trends of the non-economic effect and the cumulative variable through first differencing. Both results, as reported in Columns (3) and (4) in Table 4, show that the non-economic effect is negatively correlated with past purchases, suggesting the presence of hedonic decline with past purchases.

5.4.3 Habit Formation

As a subscription program may encourage customers to purchase upon joining, this increase in the short run may also lead to a habitual increase in the long run (e.g., Wood and Neal 2009). This mechanism would predict that a high level of customer purchases in the long term is a result of state dependence (i.e., habits formed) based on the increase in purchases in the short term. However, as we find that the non-economic effect is negatively correlated with past purchases, we conclude that habit formation is unlikely to be the driver of the purchase patterns.

5.4.4 Status

Extant research on loyalty programs show that members may 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). This literature 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, the non-economic effect of the program should be smaller as the number of members increases. To test this hypothesis, we again estimate Equation (4) and Equation (5), using the total number of members as the explanatory variable. Column (5) and Column (6) in Table 4 report the results from these models. While results from the linear model indicate that the non-economic effect of the program is decreasing with the number of members, the correlation becomes insignificant when first differencing is done. We conclude that status effect is weak in the context of subscription programs.[30]

So far, we have examined the four mechanisms in isolation. We now explore possible co-existence of multiple mechanisms using a unified regression framework because we have estimated the non-economic effect at the customer-month level. The last column in Table 4 reports the results from the first difference regression model where all three moderators are included in the vector X_{it}. We find the insights from the above analyses continue to hold. The purchase pattern is consistent with the hypothesis that members increased their purchase due to positive affect towards the firm. After accounting for all the other patterns, we also find evidence that customers exhibited a sunk cost fallacy. This finding is in line with past work that documents consumer response to sunk costs in various contexts. For instance, Arkes and Blumer (1985) show that in the context of theater season tickets, the cost of season tickets can affect

[30] To tease out the status effect, we have also examined the non-economic effects for different cohorts. We do not find a systematic pattern. The results are available from the authors upon request.

show attendance. People who pay full price for the season tickets attend more shows than those who receive discounts. Ho et al. (2017) find that the price of the license plates has influence on car usage. Given the sensitivity to sunk costs, it may benefit companies to make the initial payment more salient.

6. Robustness Checks

In this section, we analyze the robustness of our findings by addressing potential confounding effects of self-selection and unobservables, different treated groups, and different outcomes of purchase behavior.[31]

6.1 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, i.e., the treatment status is not correlated with the unobservables.

While the unconfoundedness assumption is usually not directly testable, we present an additional piece of evidence to alleviate the concern for this assumption. Specifically, we use late adopters, rather than non-members, as controls for early adopters. The late adopters could display a closer resemblance to 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

[31] We estimated both the treatment and non-economic effects 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 report the effect on purchase amount based on all months post subscription. Other results are available from the authors upon request.

November 2016 as the control group, allowing us to have enough customers to match from and enough time periods to estimate the effect. We find qualitatively similar results. On average, customers increased monthly purchase amount by $31.38 (std. err. = 2.18). After accounting for the economic benefits, their monthly purchase amount increased by $18.56 (std. err. = 1.92).

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, in Tables 5a and 5b, we replicate the analysis for members who joined the program during other months. We also measure the average treatment and non-economic effects across cohorts by estimating the DD model in Equation (1) on a matched sample combining all the (weighted) samples from the cohort-specific analyses. The last row in Tables 5a and 5b reports the average treatment and non-economic effects across cohorts. Our results suggest that the effects of subscription on purchase amount across several cohorts are largely similar and our results are robust.

Inset Tables 5a and 5b about here

6.3 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 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 the purchases combined between online and offline channels and retaining the operationalization of all covariates in Table 2. We find the treatment and non-economic effects on total purchase amount is \$26.72 (std. err. = 1.68) and \$19.41 (std. err. = 1.75), respectively, suggesting that in our context, the online and offline channels are only weak substitutes.[32] In summary, the program is effective in lifting overall revenue for the firm in our context.

7. Profitability of the Program

A successful subscription program can lead to an increase in sales but may induce additional costs to the firm. As the profitability of the program will depend on whether the increase in revenue outweighs the additional cost, we provide the profitability of the program. We also investigate the characteristics of members who were most profitable. We leverage information on how members utilized program benefits and construct a cost measure to calculate the profitability. We note that some of the cost information for the profit calculation are not readily available (e.g., profit margins at the product level). Therefore, the cost measure we construct in this section is a proxy for the actual costs to the firm. Our intention is to illustrate the key tradeoffs a firm may face when launching or expanding the scale of its subscription program.

We construct the gross profit as the increase in revenue less the increase in the cost of goods sold (COGS).[33] We also consider three types of costs associated with the program: free shipping service, \$3 gift card per month, and free samples with purchase. Based on our results, members ordered more frequently than non-members, resulting in slightly higher shipping costs for the firm. There were costs from gift cards and free samples as well. Using data on how members utilized the benefits of the program, we construct the cost measure as the increase in

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

[33] We observe that COGS typically ranges from 20% to 30% of revenue in the beauty industry. In our analysis, we assumed that COGS is 25% of the list price.

shipping cost (number of orders made multiplied by the shipping cost per order) and add the costs incurred through redemptions of gift cards and free samples.[34]

We find that the program generated an average monthly net profit of $13.71 per member. While the information on customer purchases beyond the first year of subscription rollout is not available, we also calculate the lifetime profit based on our estimates of the one-year treatment effects and the information on member retention. The expected lifetime profit per member is $366.[35] We find there is significant variation in profit across members. Figure 4 plots each member based on the cost incurred through subscription and (gross) profit.

Insert Figure 4 about here

Taking a step further, we apply the K-means clustering algorithm to segment members based on the two measures. We find members were grouped into three segments, one small segment (segments 1 accounting for 14% of members) and two large segments (segments 2 and 3 accounting for 46% and 40% of members, respectively). Segment 1 (empty diamond in Figure 4) includes the members who contributed highest profit but incurred largest cost for the firm. This segment generated about 30% of the total profit. The largest segment, segment 2 (black square in Figure 4) represents the members who generated a sizeable increase in profit but did not incur as much costs related with subscription. Segment 3 (grey square in Figure 4) is similar to segment 1 in terms of the incurred cost but differs significantly in terms of the profit.

We also find the three segments differed in terms of their observed characteristics. Table 6 reports the summary statistics of observed characteristics for the three segments. Interestingly,

[34] We cannot reveal shipping costs which the firm paid to delivery services companies due to the non-disclosure agreement and note that shipping costs in Asia were considerably lower than those in the US. The cost of a sample was derived from the price of the full-sized regular product assuming that cost is proportional to size.

[35] We observe 55% of the members renewed their subscription after the first year. Assuming that the program's attrition rate stayed constant over time, the expected lifetime profit per member is: Profit per year / (1 – Renewal rate) = $13.71 × 12/(1-0.55) = $366.

the members who contributed highest profit for the firm upon subscription (segment 1) are not the ones who purchased most prior to subscription. Rather, they were relatively less active customers in the pre-subscription period but were willing to explore new products (exploratory), had repeated purchases (repetitive), and responsive to promotions (promotional) so that they re-engaged with the firm post subscription. The high-value customers (segment 3) based on past purchases, while enjoying program benefits, increased their purchases only moderately, which is likely due to a ceiling effect. These results can assist managers in scoring and targeting future customers for the subscription program.

Insert Table 6 about here

8. Conclusions

In this paper, we examine the causal effect of a subscription 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 effect on purchase amount is attributed to the economic benefits of the program. The effect is economically significant, persistent over time, and heterogeneous across customers. Our main findings are robust to potential confounding effects of self-selection and unobservables, different treated groups, and different outcomes of purchase behavior.

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

Our results shed light on the practice of subscription-based businesses for customer retention and development. In our context, we find subscription programs are broadly effective in lifting sales and enhancing customer engagement (e.g., breadth and depth of purchases in products and categories). While we study one subscription program launched by a single firm, the insight that the effect of subscription goes beyond the economic benefits of the program 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). As members' enhanced status encourages purchases, managers can emphasize the exclusiveness of the program to boost sales. 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 the first attempt to identify the effect of subscriptions in the retail area on customer behavior, naturally there are limitations that should be acknowledged and addressed in future research. First, as our study focused on the effect of a subscription program in a given firm, 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. The subscription program examined in this research is an online-only program, while members in other subscription programs could benefit at both online and offline channels (e.g., Amazon, Barnes & Noble). With that in mind, we hope our approach provides a framework for further studies on subscription programs with various pricing schemes or framing (e.g., subscription programs with more varying time windows) in other product categories and across channels. Second, while we examined customer behavior in the first year of subscription rollout, it could be interesting to measure the long-term (e.g., 3 years) effect of subscriptions on customer behavior beyond the first year of subscription rollout. Finally, 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. 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," *Annals of Statistics*, 47 (2), 1148-1178.

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

Bertini, Marco and Luc Wathieu (2008), "Research Note—Attention Arousal through Price Partitioning," *Marketing Science*, 27 (2), 236-246.

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," *Econometrics Journal*, 21 (1), C1-C68.

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.

Gopalakrishnan, Arun, Zhenling Jiang, Yulia Nevskaya, and Raphael Thomadsen (2021), “Can Non-tiered Customer Loyalty Programs Be Profitable?,” *Marketing Science*, 40 (3), 508-526.

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.

Hamilton, James Douglas (2020), *Time Series Analysis*, Princeton University Press.

Hanushek, Eric A. (1974), “Efficient Estimators for Regressing Regression Coefficients,” *The American Statistician*, 28 (2), 66-67.

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.

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.

Iyengar, Raghuram, Kamel Jedidi, Skander Essegaier, and Peter J. Danaher (2011), "The Impact of Tariff Structure on Customer Retention, Usage, and Profitability of Access Services," *Marketing Science*, 30 (5), 820-836.

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.

Lewis, Michael (2006), "The Effect of Shipping Fees on Customer Acquisition, Customer Retention, and Purchase Quantities," *Journal of Retailing*, 82 (1), 13–23.

Lewis, Michael, Vishal Singh, and Scott Fay (2006), "An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions," *Marketing Science*, 25 (1), 51–64.

Liang, Kung-Yee and Scott L. Zeger (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73 (1), 13–22.

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.

Sato, Kazuo (1972), “Additive Utility Functions with Double-Log Consumer Demand Functions,” *Journal of Political Economy*, 80 (1), 102–124.

Tellis, Gerard J. (1988), “The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales,” *Journal of Marketing Research*, 25 (4), 331-341.

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 versus 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
Purchase characteristics			
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 price 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
Address	Coordinates of home address		

Table 3a: 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: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 3b: Non-economic 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: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 4: Regression Results for the Non-economic Effects

	(1) Linear	(2) First Diff.	(3) Linear	(4) First Diff.	(5) Linear	(6) First Diff.	(7) First Diff.
Cumulative benefits ($)	-0.128***	-0.226***					-0.179*
	(0.028)	(0.023)					(0.103)
Cumulative purchases ($)			-0.019***	-0.034***			-0.009
			(0.004)	(0.003)			(0.013)
Number of members					-0.003***	-0.0002	0.0004**
					(0.0001)	(0.0001)	(0.0002)
Observations	6,710	6,710	6,710	6,710	6,710	6,710	6,710
R-squared	0.088	0.083	0.085	0.081	0.070	0.008	0.084

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors clustered at the individual level appear in parentheses.

Table 5a: 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***	40.57***	28.95***	25.53***
	(1.75)	(4.70)	(2.33)	(2.03)
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)
Average	29.35***	38.77***	29.61***	28.29***
	(2.24)	(3.64)	(3.05)	(2.87)

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 5b: Non-economic Effects across Cohorts

Cohort	All months	Month 2	Months 3-4	Months 5+
Feb. 2016	20.29***	33.93***	29.22***	17.22***
	(2.05)	(1.29)	(4.05)	(1.59)
Mar. 2016	17.67***	32.87***	25.69***	15.03***
	(2.14)	(1.38)	(2.91)	(1.75)
Apr. 2016	17.91***	37.28***	22.26***	12.31***
	(1.96)	(9.51)	(2.70)	(1.99)
May. 2016	17.23***	32.24***	15.02***	15.28***
	(1.91)	(1.12)	(2.42)	(1.64)
Jun. 2016	23.50***	36.94***	30.64***	19.02***
	(2.14)	(3.21)	(2.71)	(1.86)
Jul. 2016	19.51***	29.03***	21.87***	16.44***
	(2.03)	(4.18)	(2.60)	(1.80)
Aug. 2016	24.99***	34.22***	28.77***	22.02***
	(2.93)	(1.35)	(3.10)	(2.69)
Average	20.09***	34.63***	24.67***	17.19***
	(2.35)	(1.26)	(2.98)	(2.21)

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table 6: Profitability and Characteristics by Segment

	Segment 1	Segment 2	Segment 3
Profit ($)	35.21	12.53	16.16
Cost ($)	5.84	1.52	4.70
Segment size (% of all members)	14%	46%	40%
Customer-firm relationship			
Tenure	6.46	6.02	6.37
Breadth	1.50	1.38	2.66
Depth	1.57	1.29	2.45
Basket size ($)	13.92	26.94	29.54
Monthly purchase amount ($): Category-level	1.56	2.45	3.53
Monthly purchase amount ($): Std. Dev.	14.22	18.79	25.71
Purchase characteristics			
Exploratory	53.09	4.03	23.41
Repetitive	0.35	0.02	0.15
Promotional	0.63	0.58	0.36
Demographics			
Age	36.02	35.37	35.80
Gender	0.98	0.93	0.95

Figure 1: Customer Purchases of Members versus Non-members

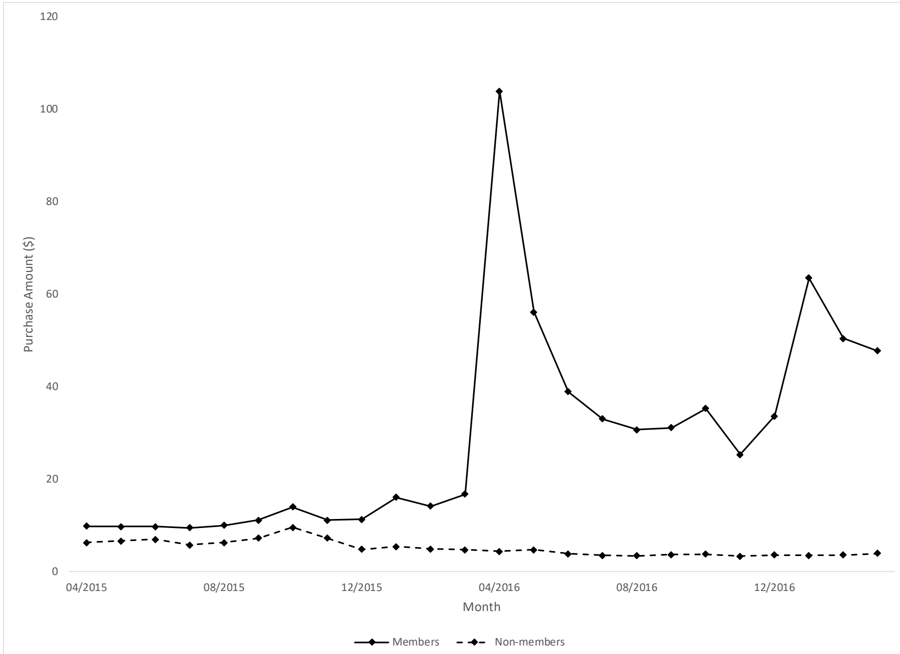
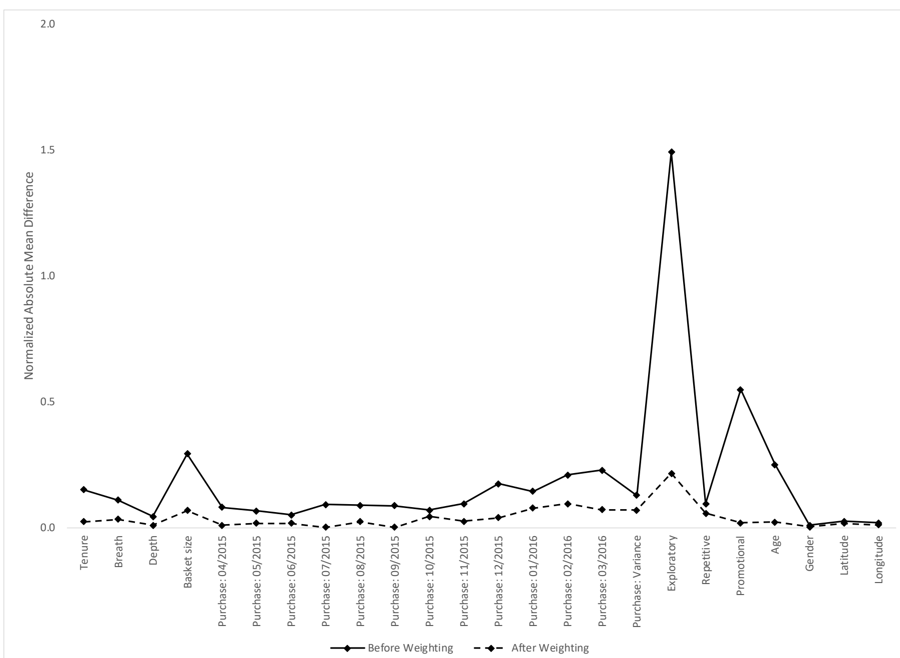


Figure 2: Covariate Balance Before and After Weighting



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 Treatment Effect

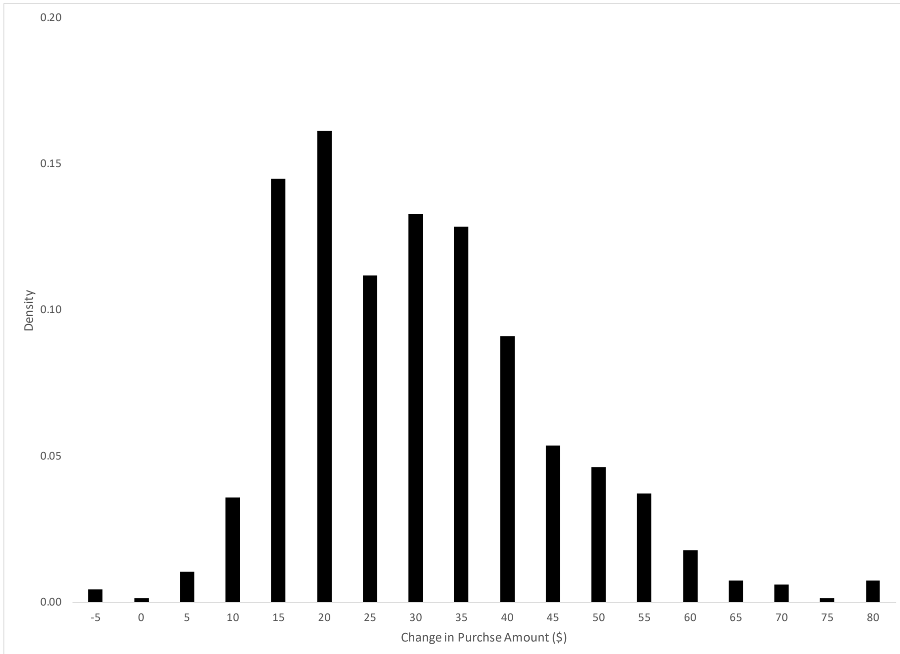


Figure 3b: Distribution of the Non-economic Effect

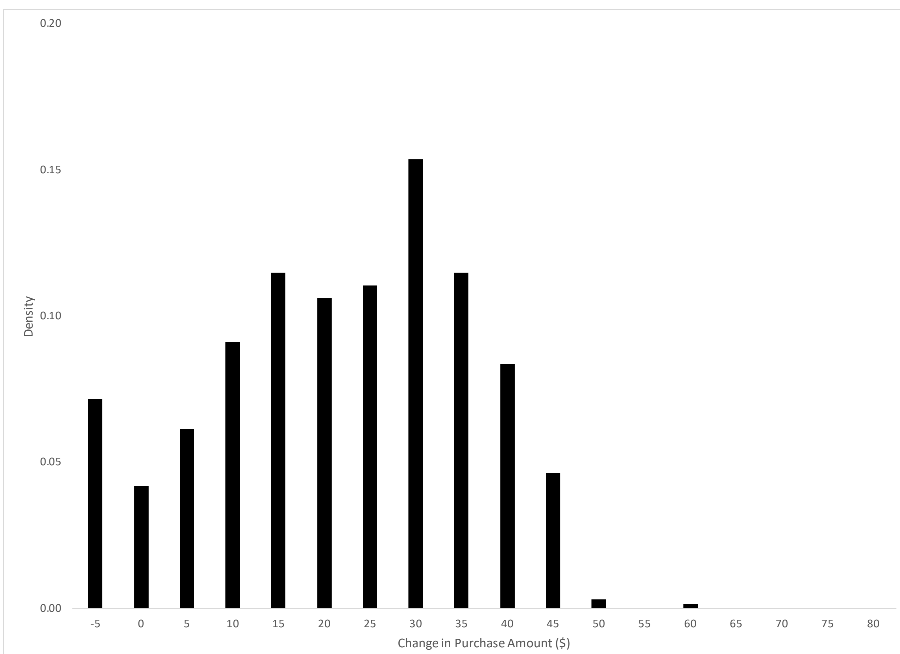
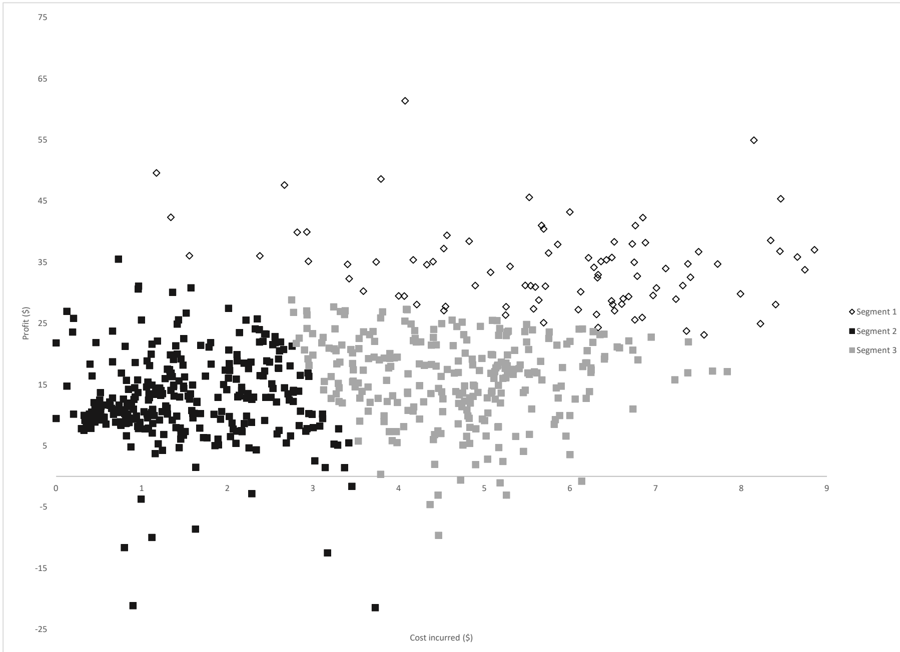


Figure 4: Scatter Plot of Profit and Cost



Appendix A: Placebo Tests

We defined a placebo treatment on members six months prior to when the actual subscription took place and estimated the effect of such a placebo treatment on customer purchases using a DD model on all weighted pre-subscription data. If the parallel trend assumption holds, we should expect null effects. Table A1 reports the results from the placebo tests. These results indicate that the members and the matched non-members have statistically comparable pre-treatment time trends.

Table A1: Placebo Tests

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

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Placebo tests on the variety metrics were left out. By design, purchase amount of known products (and categories) equals to the monthly purchase amount before subscription and the purchase amount of new products (and categories) remains unchanged (at zero) before subscription.

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

Tables B1 and B2 report the average treatment and non-economic effects of subscription on customer purchases, respectively, where the outcome measure is transformed with natural log.

Table B1: Treatment Effects 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: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Table B2: Non-economic Effects 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: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust standard errors appear in parentheses. Customer purchases in the month of adoption (month 1) were excluded in the analysis.

Web Appendix A: Log-linear Demand Model

In Web Appendix A, we show that our demand model is consistent with utility maximization under Cobb-Douglas preference (e.g., Sato 1972).

Consider a customer with a budget I and consumes Y units of products in the focal company and Z units of the outside product. The price for the inside product is given by p and the price for the outside product is normalized to 1.

The customer has Cobb-Douglas preference, that is, her utility from (Y, Z) is given by:

$$U = Y^{\alpha} Z^{1-\alpha}. \tag{WA1}$$

And the customer has the following budget constraint:

$$pY + Z = I. \tag{WA2}$$

Then the utility-maximizing customer's demand (in the focal company) is as follows:

$$\log(Y) = \log(\alpha) + \log(I) - \log(p). \tag{WA3}$$

Given (WA3), if there were an increase in the customer's budget by ΔI, the resulting demand function would be as follows:

$$\log(Y) = \log(\alpha) + \log(I + \Delta I) - \log(p).$$

Thus, any changes in the budget are included within the logarithmic term. By doing so, the resulting demand model remains consistent with the underlying utility maximization. In our context, the increase in the budget for members was due to the monthly gift card. With that being in the case, the amount of gift card is included within the logarithmic term.

To work with empirical data, we relaxed the constraint that the income and price elasticities are unity. Demand functions of similar form have also been used to study consumer demand (e.g., Berndt et al. 1977, Young 2012, 2013, Aguiar and Bils 2015) and the demand for water and land (e.g., Gaudin et al. 2001, Griffin and Chang 1991). We also generalized the model to use outcome variables other than log quantity as the dependent variable. In our main analysis, for the ease of interpretation, we used the level of demand as the dependent variable. Finally, we included customer and month fixed effects and estimated our empirical model.

Web Appendix B: Operationalization of Store-level Price

In the main analysis, we proxied the customer-level prices using group-level (members and non-members) prices. We did so to provide appropriate information on the price discounts due to the subscription program and its variation on a monthly basis. Specifically, we first defined the group-level product price, p_{gjt}, as the minimum price of product j paid by group g (members or non-members) in month t. The group-level store price is then calculated as the weighted average of the product prices where the weights are based on the quantity purchased by members. The store-level price ($Price_{it}$) for customer i in month t is:

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

where Q_{jt} is the quantity of product j purchased by members and J_t is the set of products that were purchased by both members and non-members in month t.

With this operationalization, we captured the upper bound of the discounts offered to members as the products which members received deeper discounts for were purchased more often by them and are weighted more so in our formulation. By doing so, we obtained a conservative estimate of the non-economic effect of subscription.

To evaluate the robustness of our results, we replicated the main analyses using a few alternative prices. We defined 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 (19.81).

We also tested an 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 constructed a basket of products that include all the products offered by the website and imputed missing prices with listed prices. We find the price variation across members and non-members was masked by the noise through the imputation of price information, probably due to a large number of products offered at the website and sparse price information in our e-commerce context. Our choice of the basket of products in the main analysis (those products purchased by both members and non-members) resulted in a store-level price that captured the price variation across members and non-members well while ensuring a sufficient number of products are included for the analysis.

We also tested alternative definitions for the weights. For example, we defined the weights as quantity based on purchases by both members and non-members. The non-economic effect estimates are qualitatively similar when alternative weights are specified (22.56).

Finally, we specified an alternative price variable at the individual level. We used a two-step procedure following Thomassen et al. (2017):

Step 1: Obtain group-level product prices as the minimum price paid by the group and construct group-level category prices using weighted-average product prices where the weights are total quantity purchased by both members and non-members, i.e., $P_{gct} = \frac{\sum_{j \in J_{ct}} P_{gjt} \cdot Q_{jt}}{\sum_{j \in J_{ct}} Q_{jt}}$, where P_{gjt} is the minimum price of product j paid by group g (members or non-members) in month t, Q_{jt} is the quantity of product j purchased by both groups, and J_{ct} is the set of products in category c purchased by both groups in month t.

Step 2: Based on category-level prices, construct store-level prices using individual-specific weights. The weights are chosen to be an individual's purchase shares of the categories throughout the data (two year) period, i.e., $P_{it} = \frac{\sum_c P_{gct} \cdot Q_{ic}}{\sum_c Q_{ic}}$. The resulting price variable varies at the individual-month level and reflects each customer's idiosyncratic preference for the product categories.

The estimated non-economic effects (23.31) is comparable to the one reported in the manuscript.

Web Appendix C: Identification of the DD Model with Covariates

We used a simulation study to understand how different patterns in purchases 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 simulated the purchase patterns of two members with different levels of baseline spend and one non-member under various sets of parameters. Specifically, we chose 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 used the values of the simulation parameters to mimic the value found in our real data. The variation in prices (between the two groups and over time) mimicked our data while the levels of prices were 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 WC1 to WC4 show the purchase patterns over a 24-month period, excluding the month of adoption (April 2016), under different sets of parameters. Figure WC1 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 there is no non-economic effect of the program, as expected, purchase patterns are constant at the level of baseline spend and do not change over the data period. Figure WC2 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 there is no non-economic effect, 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 across 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 WC3 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 there is no non-economic effect. 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 WC4 plots the case when $\tau = 18$, $\beta_1 = 0$, $\beta_2 = 0$. A positive non-economic 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.

Figure WC1: Customer Purchases When $\tau = \beta_1 = \beta_2 = 0$

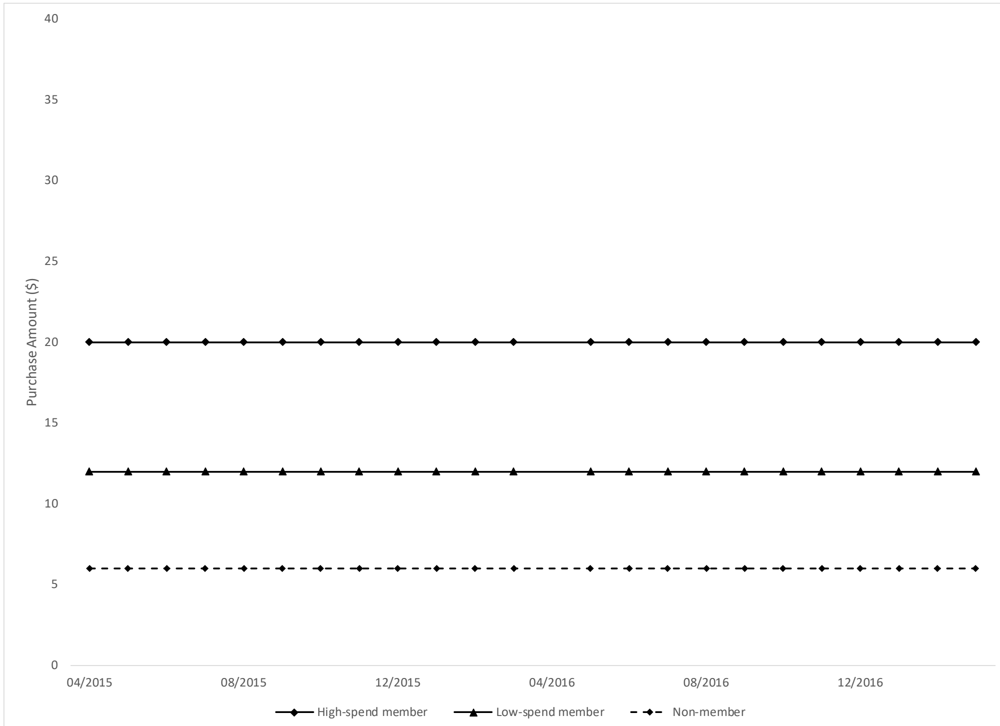


Figure WC2: Customer Purchases When $\tau = 0, \beta_1 = -3, \beta_2 = 0$

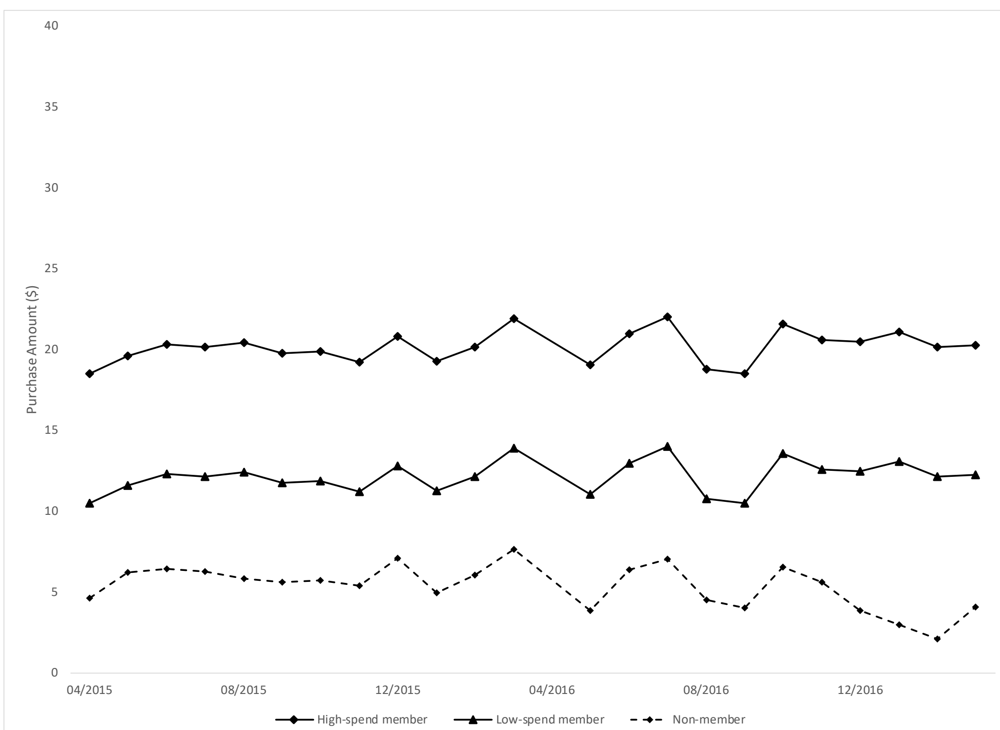


Figure WC3: Customer Purchases When $\tau = 0, \beta_1 = 0, \beta_2 = 3$

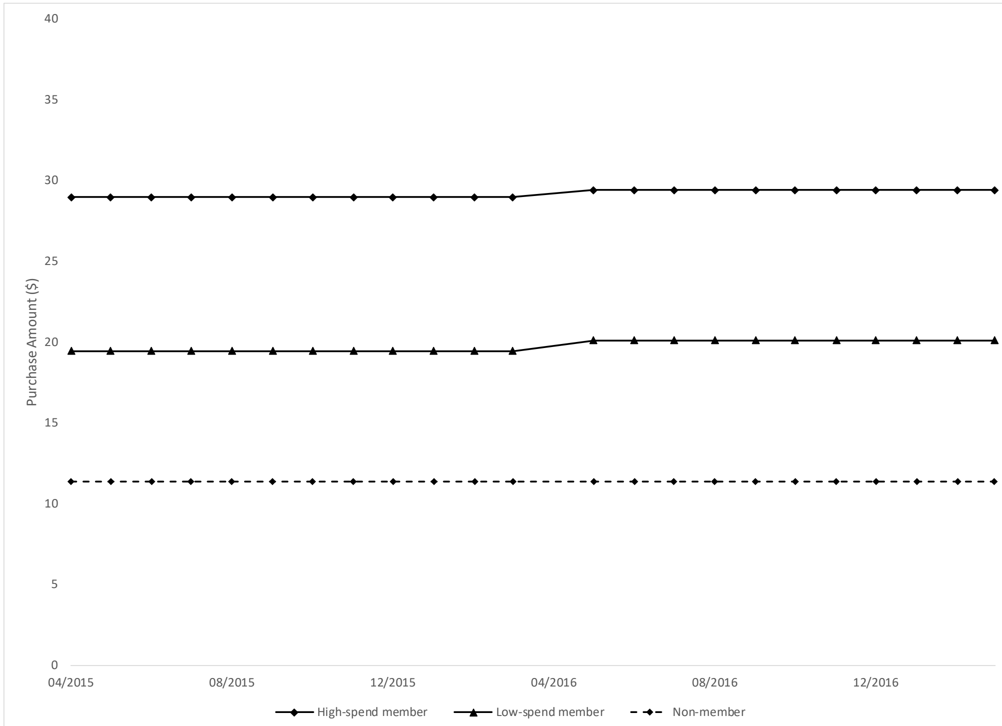
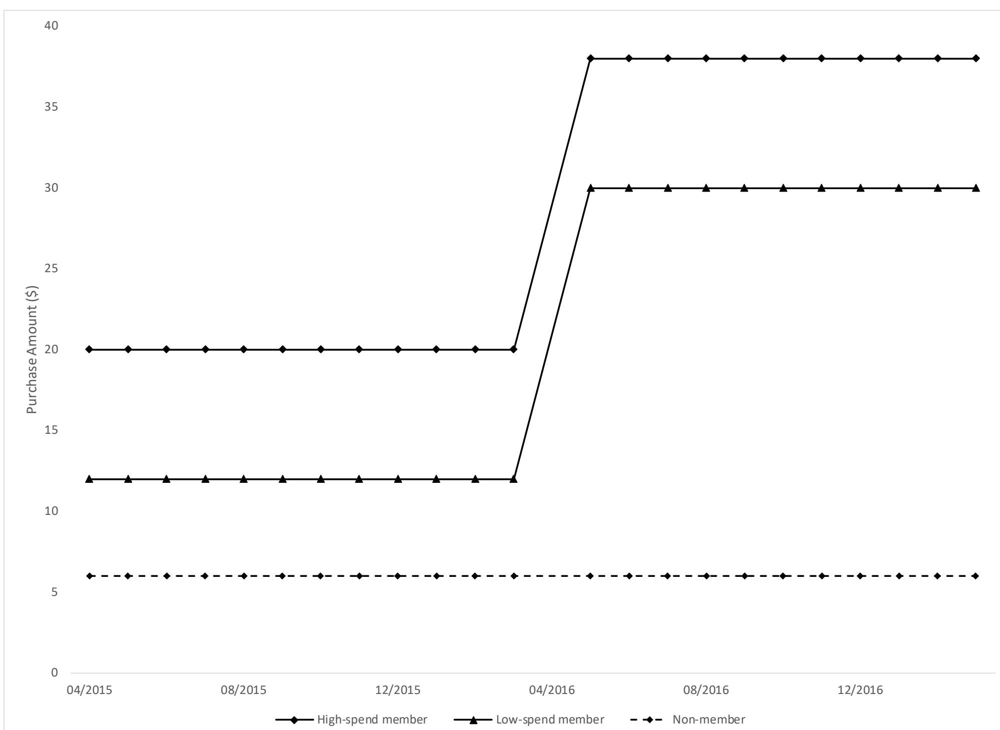


Figure WC4: Customer Purchases When $\tau = 18, \beta_1 = 0, \beta_2 = 0$



Web Appendix D: Estimation

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

WD1. Implementation Details for GRF

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-processed the data. We next built a random forest that defines the weighted set of neighboring customers for each member. Finally, the treatment effects were obtained by estimating Equation (2) on the weighted sets. Below we describe each step in detail.

To accommodate the fixed effects, we first demeaned 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 implemented an orthogonalization procedure: we residualized the (demeaned) variables by the covariates x using separately trained regression forests. The final GRF is trained on the residuals instead of the original variables.

To build the random forest, we grew a decision tree through iteratively partitioning the data into subgroups (leaves). In each iteration, the algorithm chooses a 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 $\hat{\theta}_{C_1}$ and $\hat{\theta}_{C_2}$ 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 of the estimators from Equation (2) and 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 chose by cross-validation. Finally, in the context of panel data, we accounted 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. 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 $w_i(x)$ 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 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. Y_{it}, M_{it}, $\log(P_{it})$, and $\log(B_i + G_{it})$ are residualized by the fixed effects and covariates x.

We used 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.

We chose the hyper-parameters of the model by cross-validation. We optimized on the three core hyper-parameters: node size (stopping criterion), sample fraction, and number of splitting covariates in a tree. For node size, we considered 1, 2, 3, 4, 5, 10, 15, and 20. For sample fraction, we considered 0.1, 0.2, 0.3, 0.4, and 0.5. For number of splitting covariates, we considered $\sqrt{n}$, $2\sqrt{n}$, $\frac{n}{3}$, $\frac{n}{2}$, where n is the number of covariates (n = 72). We used cross validation to choose a set of hyper-parameters that gives the smallest out-of-bag error.

WD2. Doubly Robust Estimator

In order to obtain the treatment effect across members and examine the average effectiveness of the program, we constructed 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, $i = 1, \ldots, n$, $\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}(X_i)$. 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 E: 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. 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 used 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 simulated 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 allowed the model parameters to be non-linear functions of the covariates. The fixed effects were set to be zero. Finally, the error terms were drawn from the standard normal distribution to generate the outcomes.

We applied our proposed estimation method to estimate heterogeneous treatment effects. Figures WE1 to WE3 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 WE1: Parameter Recovery of τ

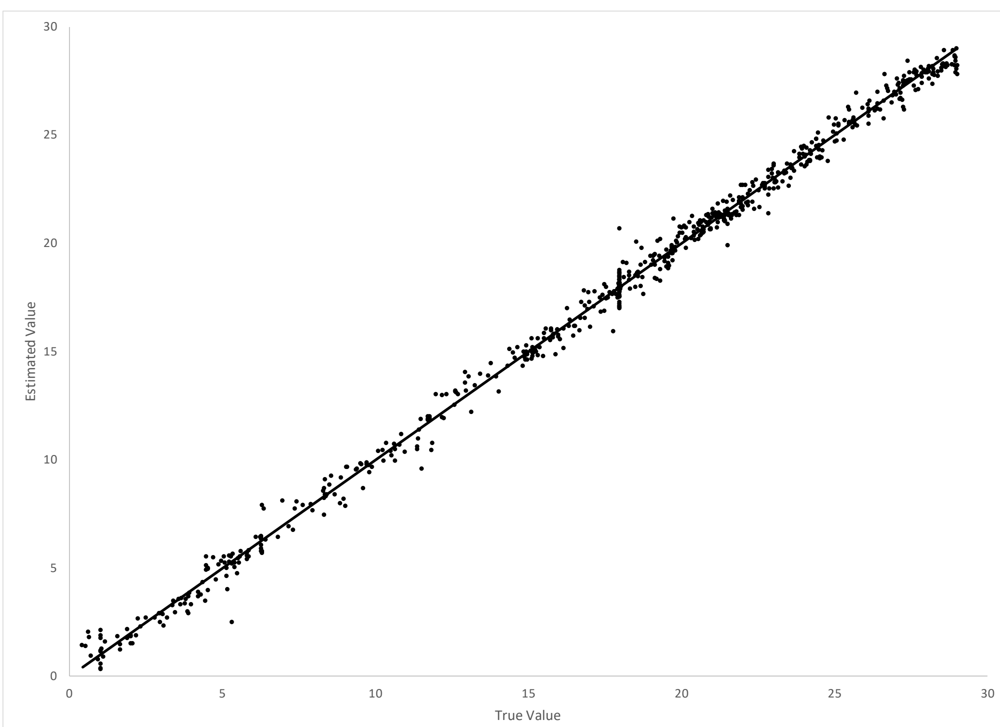


Figure WE2: Parameter Recovery of $\boldsymbol{\beta}_1$

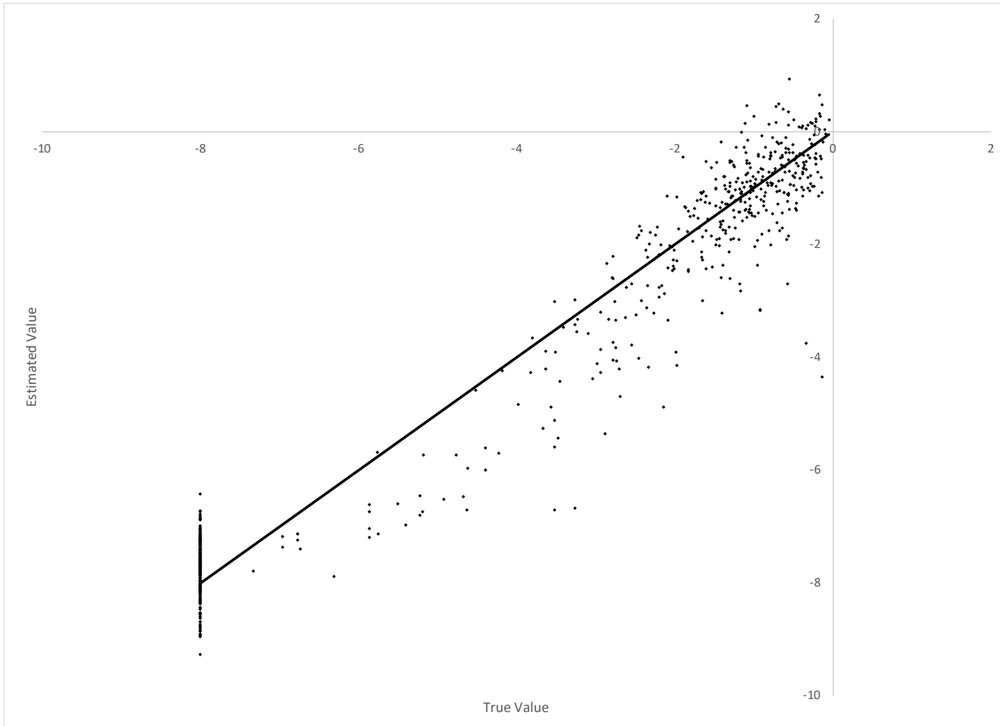
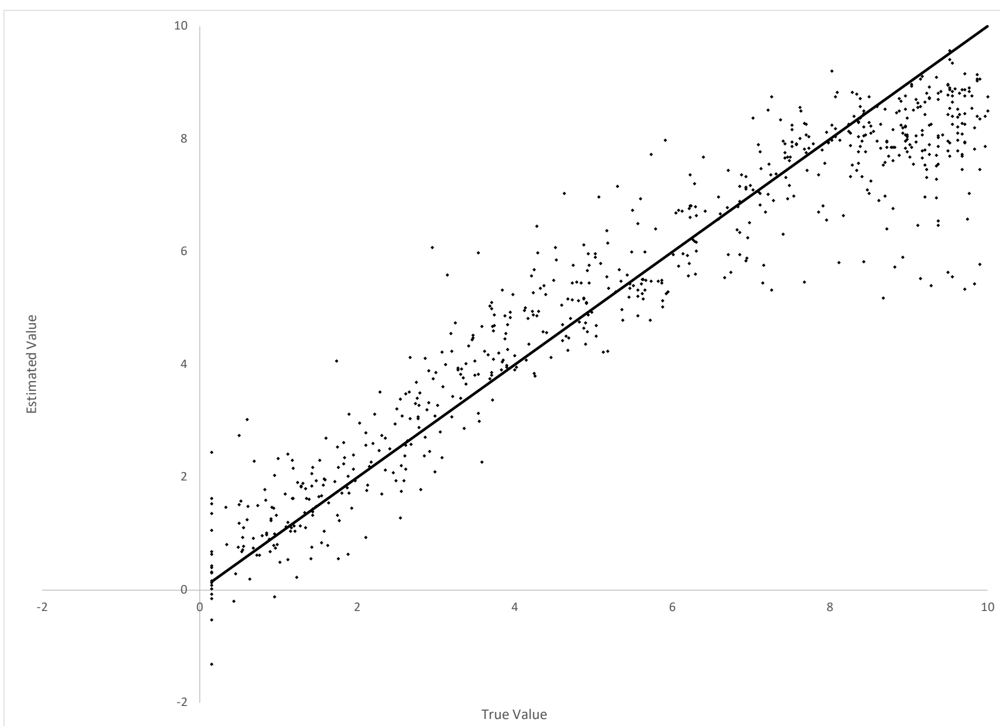


Figure WE3: Parameter Recovery of $\boldsymbol{\beta}_2$



References

Aguiar, Mark and Mark Bils (2015), "Has Consumption Inequality Mirrored Income Inequality?," *American Economic Review*, 105 (9), 2725-2756.

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

Berndt, Ernst R., Masako N. Darrough, and W. Erwin Diewert (1977), "Flexible Functional Forms and Expenditure Distributions: An Application to Canadian Consumer Demand Functions," *International Economic Review*, 18 (3), 651-675.

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," *Econometrics Journal*, 21 (1), C1-C68.

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.

Gaudin, Sylvestre, Ronald C. Griffin, and Robin C. Sickles (2001), "Demand Specification for Municipal Water Management: Evaluation of the Stone-Geary Form," *Land Economics*, 77 (3), 399-422.

Griffin, Ronald C. and Chan Chang (1991), "Seasonality in Community Water Demand," *Western Journal of Agricultural Economics*, 207-217.

Sato, Kazuo (1972), "Additive Utility Functions with Double-Log Consumer Demand Functions," *Journal of Political Economy*, 80 (1), 102-124.

Thomassen, Øyvind, Howard Smith, Stephan Seiler, and Pasquale Schiraldi (2017), "Multi-Category Competition and Market Power: A Model of Supermarket Pricing," *American Economic Review*, 107 (8), 2308-2351.

Young, Alwyn (2012), "The African Growth Miracle," *Journal of Political Economy*, 120 (4), 696-739.

Young, Alwyn (2013), "Inequality, the Urban-Rural Gap, and Migration," *Quarterly Journal of Economics*, 128 (4), 1727-1785.