TWO ESSAYS ON PRODUCT SUBSCRIPTIONS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

YILONG LIANG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

TONY HT CUI, ADVISOR

May, 2021

Acknowledgements

I knew getting a Ph.D. would be hard, but I never expected it to be this hard. I would never finish this long journey without the help of many people.

I want to thank my advisor Professor Tony Cui for leading me into this exciting field of marketing. I am also grateful for his advice through my entire process pursuing the doctoral degree. I am also indebted to Professor George John, who always provide the most straightforward, sometimes ruthless comments, about my thoughts. I did benefit a lot, both in terms of knowledge and mental toughness, from the many long disscussions with him. I appreciate the great help that Professor Mark Bergen has given me. He is one of the nicest person I have ever seen and he will always be my role model. I also want to thank Professor Amil Petrin for serving on my committee and providing constructive advice and comments for my dissertation.

I would also like to extend my gratitude to all the other faculties in the marketing department. Marketing is such a diverse field and they have shown me the many different facinating aspects of this field.

I will never forget my fellow Ph.D. students in the department, Ruitong Wang, JJ Chen, Ryan Wang, Maria Rodas, Lei Zhuang and Yuanchen Su. They provided a lot of academic and emotional support for me.

Finally, I want to acknowledge the help from Sandy Herzan and Deb Basarich. Their administrative support has saved me many trouble.

Dedication

To my granny, who I guess would be very happy, had she lived to see the day when I become a doctoral student.

Abstract

Product subscription, i.e. the business model that customers pay to get periodical delivery of certain products, is trendy in recent years. However, given the rising popularity, researches are still lagging behind. In my dissertation, I explore the effects of subscription products on both other not-for-subscription products and the products that are available for subscriptions in a multi-product context. By collecting data from a grocery retailer which rolls out subscription plans for several of its products, I investigate how subscriptions affect customers' purchases of different products. I find that subscriptions increase customers' purchases of the other products, which is partly due to the reminding effects of subscriptions during product subscriptions' delivery. The availabilities of subscription options also facilitate customers' purchases and increase the overall sales of the products that are available for subscription. Although the overall effects are positive, they are heterogeneous and retailers should be cautious about some of the possible negative consequences.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Product subscriptions refer to the business model that customers pay to get access to certain tangible products on a recurring basis. This emerging business model is gaining popularity and more and more products are available for subscription. Although subscription business model is not new and has been employed in the service domains for a long period of time, experience and insights from service subscriptions are not completely applicable for product subscriptions given the differences between tangible products and intangible services. The two subscription models also differs fundamentally in the way they are operated. For example, in service subscription, customers usually have unlimited access to the service once they subscribe due to the negligible marginal cost of services, e.g. software. However, customers in product subscriptions are always restricted by the quantity and delivery interval specified for the subscription plans. There are subtle distinctions between service subscriptions and product subscriptions, and product subscriptions are worth their own scrutiny.

In my dissertation, I investigate the effects of product subscriptions in a multi-product retailing context. I conduct empirical researches by using a unique dataset containing customers' shopping transactions at a grocery chain in a period when subscriptions of some products become available. Such an exogenous shock gives me a great opportunity to examine the effects of subscriptions. In a multi-product context, the study for the effects of marketing interventions on certain products/categories should not be limited to the focal products/categories per se. Thus, I examine both the effects of subscription on other products (those that are not available for subscription) and the focal products (those that are available for subscription) themselves. My dissertation contains primarily two essays that focus on these two different aspects. There are basically two stages for product subscriptions: first, the retailer has to introduce subscription options to make them available for the customers, and then, customers can choose whether to subscribe to certain products. There are some critical differences between these two stages. Table A.1 illustrates the relationship between my two essays.

[Insert Table A.1 about here]

In Essay I, I focus on the effects of subscription on other products. In a grocery retailing context where only a small portion of products are available for subscription, the effects on other products are more substantial. Because the introduction of subscription is available to all customers, there is not a group of customers who are unaffected by the introduction of subscription plans. As a result, I do not have a control group for studying the effects of subscription introduction. However, I do observe the behaviors of customers who subscribe and those who do not. Thus, I have to take advantage of the quasi-experimental setting of the context and use the customers who never subscribe as a comparison group to explore how actual subscribing to products affect customers' purchases of the other products.

Subscribing to a product should also have a greater impact on customers than the mere exposure to the subscription options. It is hard to envision a scenario where customers' purchases of tomatoes changes simply by knowing the existence of the subscription options of chicken eggs. Moreover, in our case, because the effects of subscribing to products should just be a subset of the effects of the introductions of subscription (because subscribing happens only after the introduction of subscription options), it is easier for us to identify the mechanisms for how product subscriptions affect customers' behaviors.

For product subscriptions, perhaps what worries marketers the most is that subscriptions may hurt the sales of other products. The sales drop could be due to either the substitution from the products that are available for subscription or the less shopping trips by the customers. However, our results show that on the contrary, subscriptions increase customers' purchases of the other products, although the effects of subscriptions are heterogeneous on the category level. The increase is partly because that product subscriptions function as an additional information channel: the delivery of subscription products reminds customers of the retailer and their grocery needs. I present evidence supporting this explanation: customers have higher probability of purchasing other products during the same day when subscription products deliver and the purchase of products complementary to the product that customers subscribe to increase more than those non-complementary. I also discuss the detailed implications of my findings.

In Essay 2, I examine the effects of subscription on subscription products themselves. I choose not to study how subscribing to a product affects the purchase of the product primarily due to the endogeneity issue. Moreover, it is quite intuitive that, after subscribing to chicken eggs, customers are most likely to reduce their future purchase of chicken eggs while the overall consumptions of chicken eggs may slightly rise, given customers have more purchase options to suit their needs. Thus, I emphasize on the purchase behavior changes after the introduction of the subscription plans, i.e. how customers' purchases of chicken eggs change after the subscription options become available.

The main goal for Essay 2 is to check whether and how much the purchase of a product will increase if customers are given the additional options to subscribe to the product. To understand customers' preferences between different purchase alternatives with different inter-temporal structure, I build a structural model based on the single-agent dynamic programming framework.

I estimate the structural parameters in the model using the actual transactions of customers in the period when subscription options are available. Then I simulate the scenario where subscription options are not available. The results show that subscription options are important to drive the sales of the product: if the subscription options were not available, the total quantity sold decreases by about 50%.

The two essays actually center around the key issue of the effects of product subscription. Ideally, the two questions ought to be studied under a universal framework of customers' decision making. However, it is almost impossible to write a structural model that realistically incorporates customers' decision process with subscription plans and is also tractable, as well as solvable, at the same time. Thus I choose the reduced-form approach to understand the overall effects of product subscription on customers' purchases of other products. For the effects on the focal products, I opt for a structural model that focus on the dynamics of customers' decision process because of the endogeneity issue.

My dissertation is organized as follows: Chapter 2 is my first essay studying the effects of subscription on other products; Chapter 3 is my second essay about the effects of subscription on the focal products. Chapter 4 concludes my dissertation.

Chapter 2

Essay1: How Subscriptions Affect Customers' Purchases of Other Products

2.1 Introduction

Recent years have witnessed the surge of product subscriptions. Large retailers like Amazon and Target start to employ the subscription business model (Amazon Subscribe & Save, Target Subscription Box) by providing subscription options for certain products. These gigantic retailers join the subscription business to reap a share from the rapidly growing subscription market. In the past years, the product subscription market has grown at a rate of over 100% (McKinsey 2018; Sinha, Foscht and Fung 2016). Today, we can subscribe to products in many categories, from beauty products to personal care, from home cuisine to daily clothes, from house decor to food and grocery. Customers are also welcoming these new businesses as they are now free to choose a new purchase mode and have the products delivered to their doors automatically. Over 15% of online shoppers have tried product subscription of some sort (McKinsey 2018). Although most of the product subscription businesses are yet to be proven successful, some are already among the headlines. In 2016, Dollar Shave Club, the company that periodically ships razors to customers, is sold to Unilever for one billion USD. With the change of customers' shopping behaviors and the advancement of technology, we should expect that product subscription will play a larger role in the future of the retailing industry.

The subscription model is not new and has been applied in various service domains, like telephone, gym membership, and software, for almost decades. There are some general explanations for customers' preferences for subscription. In some cases where subscription is the sole option to obtain the service, customers have no choice but subscription. The lower unit price is also an advantage for subscription. This is likely to be the case for some scenarios where retailers offer lower unit prices similar to quantity discounts when customers subscribe. However, studies show that customers have a preference for subscription over pay-per-use even when these two factors are not present. Thus, there are more profound and nuanced reasons for customers' preferences for subscription. Many of the previous researches focus on these demand-side reasons for subscription. For example, Nunes (2000); DellaVigna and Malmendier (2006); Lambrecht and Skiera (2006) document the "irrational" behaviors of customers who choose subscription contract over pay-per-use contract, though it is more expensive in terms of cost per use. They propose several possible reasons, like "insurance effect", "taxi meter effect", "convenience effect", and "overestimation effect", to explain customers' preferences for subscription contract, in the situation where subscription contract is not the "rationally" optimal contract. Another stream of researches explores the business strategies given subscription plans are present. For example, some studies investigate the pricing issues of subscription plans by comparing different structures of subscription plans (Mitchell 1978; Danaher 2002; Tian and Feinberg 2020). Some researches study the competition between subscription firms and regular firms regarding entry, pricing, companies' profit and social surplus using analytical models (Fan, Kumar and Whinston 2009; Ma and Seidmann 2015; Guo and Ma 2018). There are only some theoretical analyses about the comparisons between subscription and regular purchases (Barro and Romer 1987; Randhawa and Kumar 2008; Cachon and Feldman 2011). Surprisingly, we are not able to identify any empirical research that examines whether subscription plans should be introduced or not by a vendor in the first place. To better understand the consequences of the subscription plan, better empirical evidence and more customer insights for customers' behavior changes due to subscriptions are required.

Moreover, almost all previous researches about subscription focus exclusively on the subscription in the service domains. The critical differences between product subscription and service subscription mean that retailers like Amazon and Target cannot simply apply whatever researchers and practitioners find previously in service subscriptions to design a successful product subscription strategy. Given the growing importance of product subscription, they are worth separate attention.

For tangible products, subscriptions usually involve recurring deliveries of products for a sum paid. Although not all product subscription business models are identical, they can be characterized along roughly four dimensions: (1) delivery: multiple times; (2) delivery interval: fixed; (3) quantity per delivery: fixed; (4) billing: once for each delivery or once for multiple deliveries.[1] Given these characteristics, some of the explanations for why customers prefer subscription in the service domain are not applicable. For example, the "insurance effect" of subscription, i.e. subscribing to a certain service, e.g. phone plan, can prevent customers from paying too much when the demand is unexpectedly high, is not playing a role in product subscription because the quantity for product subscription is fixed. The main reasons for customers to choose product subscription lie in the lower monetary and time cost.[2] Subscription usually involves a larger quantity of products compared with a one-time purchase, thus the unit cost is lower. As a form of long-term contract, subscription reduces the time cost to make multiple decisions as it automates future transactions. The long-term nature also helps the customers hedge the risk of price fluctuation in the future. The traveling cost is also saved as products are usually delivered to the home under subscription. Although these reasons help with the design of a product subscription plan, direct analysis of the customers' behavior changes due to subscription plan is needed for a holistic understanding of product subscriptions. Large retailers like Amazon need more insights about how product subscriptions affect different products and categories to specially curate subscription plans that maximize their overall interests.

Subscription options could have effects on two aspects. First, the new options of subscription are likely to increase customers' overall purchases of the product (that is available for subscription) or keep them the unchanged because they are usually cheaper

[1] Examples for once per delivery: Dollar Shave Club, Blue Apron, Amazon Subscribe & Save; examples for once for multiple deliveries: magazine/newspaper subscriptions, FlowerPlus, ThriveMarket.

[2] There are other possible reasons like hope for surprises and access to exclusive products (McKinsey 2018). However, they are applicable in a relatively small context.

in terms of unit price and save customers' time. Subscriptions either cause "purchase acceleration", i.e., shifting future purchase to the current moment (Chintagunta 1993), or cause "brand switching", i.e., shifting the purchase of other products to the focal products (Gupta 1988). In both cases, customers' purchases of the focal product will not decrease.

Second, the effects on the purchases of other products that are not available for subscription, meanwhile, are less clear. If subscription functions solely as a new channel for the delivery of products, i.e. product fulfillment channel, then subscription is likely to reduce customers' purchases of the other products. The reduction could come from two different processes. First, given the relatively stable demand for products in a certain domain, e.g. grocery, if subscriptions increase the consumption of the products that are available for subscription, they are going to substitute for customers' need for other similar products ("brand switching"). Second, subscription may reduce the number of shopping trips, resulting in fewer purchases of the other products. Both researchers (Chevalier, Kashyap and Rossi 2003; Liu, Chintagunta and Zhu 2010; Zhou 2014; Thomassen et al. 2017) and practitioners are aware of the complementarities among products that customers may purchase during the same shopping trip. The use of subscription lowers the necessity of making shopping trips (both online and offline), and thus reduces the opportunity for such complementarities ("trip skipping").

In this paper, we focus on the consequences of product subscription in the multi-product context. By collecting data from a membership-only grocery retailer which rolls out subscription options for several of its product, we investigate the effect of subscription on the customers' purchases of other products that are not available for subscription. The effects of subscription on these products are less clear and are probably ignored even by the retailers. However, given that the products which are not for subscription usually take a larger share of total product sales, the effects on these products could have a more significant impact on the overall revenue of the retailers.

Our results show that subscription, on the contrary, increases customers' purchase of the other products. This suggests that subscription functions more than merely a new channel to distribute products to customers. Both complementarity and substitutability of subscription are identified on the category level. Specifically, we find that after customers subscribe to a product, the purchase of other products in the same category decreases , while the purchase of products in other categories increases, resulting in a net increase of overall purchase of other products. In other words, we do find evidence for subscription as both product fulfillment channel and other channels.

There could be multiple reasons that subscriptions increase customers' purchases of other products. One possibility is that subscription plan also serves as an informational channel. The information provision function, e.g. helping customers recognize or recall a retailer for a certain type of purchase, has been documented for other retailing elements, such as offline stores, in previous literature (Avery et al. 2012). We show the evidence for this function of subscription plans. We present the empirical evidence for subscriptions' reminding effects due to the delivery of subscription: the customers have a higher probability of purchasing other products from the retailer on the same day when the subscribed products are delivered to the customer. We further demonstrate that the reminders come partly from the delivered subscription products themselves: customers' purchases of the products that are complementary to the subscribed products increase more than those that are not.

Our research speaks directly to the trendy phenomenon of product subscription, which, so far as we know, has not been directly studied in the previous literature. The results provide direct insights for the retailers and new startups which are the potential adopters of this new business model. Our results also demonstrate a new mechanism for the complementarity between products. The mechanism we identify about how subscription increases other products' purchases shows that the complementarities between products from the same retailer could occur inter-temporally: customers' demand are stimulated when the customers are reminded of the retailer and their grocery needs during the subscription products' delivery.

2.2 Institutional Background and Data

To study the effects of subscription, we obtain data from a membership-based grocery chain in China. To purchase in this grocery chain, members are required to either log in (if purchasing in telephone, website, and app channels) or scan their member cards (if purchasing in offline stores) to use the balances in their accounts. This institution feature enables us to track individual customers' purchase history.

We collect transaction-level data from all members who make at least one purchase in the period from 10/1/2014 to 1/31/2018. In our data, there are in total more than 2,600 SKUs from 14 different categories[3] sold to customers. Some of the products are store-branded while the rest are from third-party vendors. To get a complete history of customers' purchases for our study, we use only transactions of the customers who sign up to become members before the beginning of the data collecting period, i.e. 10/1/2014. This leaves us with in total 2,175,761 transactions from 3,659 members. Each transaction contains information including order ID, member ID, order date, delivery date, SKU number, SKU name, category, amount, prices paid, order channel, order store (if purchase in offline stores). We also collect member-level information, which includes: member ID, gender, sign-up date, family size and referee information. We summarize the key statistics in Table A.2.

[Insert Table A.2 about here]

In February 2015, the grocery chain starts to roll out subscription plans for its customers. According to the management team of the grocery chain, they choose the subscription products based on customers' previous purchase patterns and customers' feedbacks. These products are usually necessities with frequent and regular needs. This exogenous change affords us the opportunity to inspect the effects of subscriptions.

The subscription plan works as follows. If customers shop in either website or app channel, they are able to find an extra purchasing option if they are searching for a product that is also available for subscription. For example, if a customer searches on the retailer's website for "COURTYARD green shell eggs" (as shown in Figure A.1), the customer can find subscription options listed alongside the options of a one-time purchase. To subscribe to a product, customers may click the option that specifies the subscription details, including price, total amount, delivery frequency, and times of delivery. There are only limited options of frequencies and number of times of delivery for each subscription option. The subscription plans are identical whether customers shop in offline stores or other online channels.

[3] This grocery chain sells 14 categories, including cleaning, gifts, fresh meats, ready to eat, seasoning, beverages, processed meat and eggs, vegetable, grain and miscellaneous, fruits, dairy, bakery and snacks, seafood, alcohol.

[Insert Figure A.1 about here]

Subscribed products are delivered on a regular basis, and most of the subscription products are delivered once/twice a week or biweekly. Unlike some other subscription models, the subscription plans offered by the focal retailer are only available for a certain period of time and for a specified number of deliveries. Customers have to re-subscribe to the product once one subscription is completed. The numbers of deliveries vary from product to product, ranging from five times to 48 times (covering roughly a year). In the period from February 2015 to January 2018, there are in total 23 SKUs from 4 different categories available for subscription. In Table A.3, we list these SKUs available for subscription together with some of the key details.

[Insert Table A.3 about here]

This retailer does offer incentives to encourage customers opt into the subscriptions. Usually, the subscription options are cheaper for 5% to 10% in terms of unit prices, depending on the frequency and times of delivery. The one-time purchase option is available for all the SKUs with subscription option, but with perhaps different unit prices. Subscriptions do not differ from regular purchases in terms of product availability.

Customers pay the lump sum price upfront for all the subscription products to be delivered (instead of paying each time when the products are delivered later). All deliveries, both for the subscription products and other regular products, are fulfilled by their courier system and delivery takes more than one day.

In our sample, there are in total 879 customers (out of the 3,659 customers overall) who subscribe at least once.

2.3 Econometric Model and Identification Strategy

Given that most subscription products are delivered on a weekly basis, we perform our analyses at the individual-week level. Our goal is to identify the causal effects of subscribing to a certain product. In our setting, the greatest challenge to identify the causal effect lies in the non-random assignments of this treatment. Consumers' subscriptions choices are self-selected and some of the customer characteristics that affect the likelihood of subscription may also correlate to the outcome variables we are interested in. In our case, as subscription requires all the money paid upfront, the customers who subscribe are more likely to have a longer relationship, higher trust, and higher brand loyalty with the grocery chain. Furthermore, some purchasing history and habits could also affect customers' subscription decisions. For example, the consumer who has previously bought a product (maybe just by chance) that is later available for subscription may be more likely to subscribe to the product after the company rolls out subscription option for the product. Some other factors like channel preference could also result in the difference of probability for subscription as the subscription options may be presented slightly differently in different channels.

Table A.4 documents the means of some key attributes in the control group and treatment group. The statistics show that customers' purchase behaviors are significantly different for customers in the two groups. In this situation, the difference-in-difference (DiD) estimator would be biased because the treated consumers who subscribe may not follow the same pre-treatment trend as the untreated consumers.

[Insert Table A.4 about here]

2.3.1 Propensity Score Matching with DiD

To address this endogeneity concern caused by self-selection, we employ a DiD model combined with quasi-experimental matching methods based on propensity scores. The main goal of using propensity score methods is to mitigate the imbalance between treatment group and control group due to self-selection. The propensity score is the conditional probability of receiving the treatment conditional on observed variables (Rosenbaum and Rubin 1983). Propensity score matching (PSM) methods match consumers in two groups and make them comparable. In this paper, we implement **three** propensity score matching algorithms to ensure the robustness of our results: static PSM (Datta, Knox and Bronnenberg 2018; Xu et al. 2017), dynamic PSM (Aral, Muchnik and Sundararajan 2009; Xu et al. 2017) and look-ahead PSM (Bapna, Umyarov and Ramaprasad 2018; Kumar, Qiu and Kumar 2018). All the matching methods can be divided into two steps: in the first step we build a prediction model to estimate the likelihood of receiving the treatment based on an array of observable covariates, and in the second step we match the treatment and control groups according to certain matching criteria.

An alternative quasi-experimental approach based on propensity score is propensity score weighting. In our case, a customer is treated only when subscribing to a product, and thus the treatment is only temporary. Most propensity score weighting methods apply the same weight to an individual and are unable to take some time-varying characteristics into consideration. Thus, we choose propensity score matching over propensity score weighting, though the former method results in possible drops of part of the sample.

Static PSM. The static PSM is the most widely used one of the propensity-score-based methods (Xu et al. 2017). In static PSM algorithm, we use the first three month's purchase data to calculate key covariates Z_i that could be used to predict the final choice of making subscriptions. Then, we run a logistic regression of customers' subscription choice on the covariates Z_i and then the propensity scores are computed using Equation 2.1. We use 1:1 match such that each customer in the treatment group is matched with at most one customer in the control group who has a similar propensity score. We use a caliper size of 0.02 times the standard deviation to control for the quality of matches.

$$\text{Propensity\_score}_i = \frac{exp[\alpha + \beta Z_i + \epsilon_i]}{1 + exp[\alpha + \beta Z_i + \epsilon_i]} \tag{2.1}$$

The key limitation of the static matching method is that it only accounts for time-invariant observable covariates Z_i which prompt the consumers to subscribe to products. The static matching algorithm fails to control the covariates that vary across time and influence consumers' decision to subscription. It will perform poorly if many variables that affect customers' subscription choices varies across time.

Dynamic PSM. Generally speaking, consumers' preferences for products change over time and are most directly affected by recent shopping behaviors. Dynamic PSM algorithm addresses this problem by allowing a treated consumer to be matched with different untreated consumers in different periods. The dynamic matching process captures time-variant covariates of purchasing behaviors that might influence the outcome of adoption of subscription. We calculate the propensity score $\widehat{\text{Propensity score}}_{it}$ regarding observed covariates Z_{it} via a logistic regression for each consumer in each period (Aral, Muchnik and Sundararajan 2009):

$$\text{Propensity\_score}_{it} = \frac{exp[\alpha + \beta Z_{it} + \epsilon_{it}]}{1 + exp[\alpha + \beta Z_{it} + \epsilon_{it}]} \tag{2.2}$$

Then, for each "customer-week" record from the treated customers, we can match it with a "customer-week" record from the control customers, which has the smallest difference in propensity scores (and the difference has to be smaller than the caliper size of 0.02 times the standard deviation). In the end, we will have a "synthesized control" for each treated customer.

Look-ahead PSM. Look-ahead PSM is the most conservative among these three methods and it suggests that we match consumers who subscribe early with the consumers who subscribe later. In look-ahead matching, we keep only the first 80 weeks of purchase data of the customers who subscribe. Then we have a sample of customers who subscribe and some who do not (but actually they will subscribe eventually). Then, we use similar methods as what we used in static PSM to calculate the covariates and compute the propensity scores. Thus, the only difference between static matching and look-ahead matching is the sample used. In look-ahead PSM, because only the customers who actually subscribe are used, many of the unobservable time-invariant characteristics that will affect customers' subscription decisions will be balanced between the treatment group and control group, which is the main advantage of look-ahead matching. For example, suppose that forward-looking consumers are more likely to subscribe a product than myopic consumers, which is hard to measure by observables. In look-ahead matching, we use only the customers who subscribe and they should be similar in terms of the degree of forward-looking. However, the disadvantage of this method is that it drops all the consumers who never subscribe in our complete data set, which accounts for a major portion of our sample.

To summarize the matching methods we use, we illustrate the ideas for three matching algorithms in Figure A.2.

[Insert Figure A.2 about here]

The key covariates Z_i or Z_{it} described above are calculated from two sources: consumers' characteristics and purchasing behaviors. Specifically, consumer characteristics include gender, sign-up date, referee information. Purchasing behaviors contain recency, frequency, monetary information (Brynjolfsson, Hu and Simester 2011), and the purchasing record of the same products that are later available for subscription. In particular, the monetary information includes the number of categories purchased, number of products purchased, total spending across different channels. The complete list of variables we use is the same as those in Table A.4. We also document the difference between the customers in the control group and the treatment group after static matching in Table A.5. The differences between the customers in the treatment group and control group are all insignificant for the variables we use for matching. The comparison between Table A.4 and Table A.5 shows the efficacy of matching in balancing the treatment and control groups.

[Insert Table A.5 about here]

Figure A.3 shows the distribution of propensity scores before and after the static matching. A graph with a more symmetric histogram between the treated and control group indicates a sample with a more comparable treatment and control group. As Figure A.3 indicates, matching methods do make the treatment and control groups more comparable [4] .

[Insert Figure A.3 about here]

After matching, we estimate the effects of subscription with the matched sample using the following DiD specification:

$$\log(Spending_{it}) = \alpha_i + \gamma_t + \beta * Subscription_{it} + \epsilon_{it} \tag{2.3}$$

In Equation 2.3, $Subscription_{it}$ is an indicator of whether a customer i subscribe to any product in week t. $Subscription_{it} = 1$ if Customer i receives subscription delivery in Week t. In other words, we define the *treatment period* to be the weeks when a customer subscribes to a product (all weeks between the first subscription delivery and the last

[4] Similar balancing effects also exist for dynamic matching and look-ahead matching. We do not present them here to reduce redundancy.

subscription delivery). Thus the customer is no longer treated after the subscription has ended. $\log(Spending_{it})$ is the log-transformed spending of not-for-subscription products of Customer i in Week t. α_i is the individual-level fixed effect of Customer i that controls for the variables that do not vary by time. γ_t is the week-level fixed effect that conotrols for the factors that affect all customers in Week t. ϵ_{it} is the idiosyncratic error term.

The key identifying assumption here is that after matching, our matched samples have similar control and treatment groups, and they have parallel pre-treatment trends of purchases for not-for-subscription products. We check this assumption using a placebo test. A placebo test uses pre-treatment data and assumes that a "fake" treatment happens in the middle point of the pre-treatment data and estimate the effect of this "fake" treatment using the same econometric model. A significant coefficient of the treatment dummy variable suggests that the treatment and control groups have non-parallel trends.

2.4 Results

In this section, we present the estimation results for the effects of subscription on other not-for-subscription products. Then we decompose the effects to see how subscriptions affect customers' purchases in different categories.

2.4.1 Main Effects of Subscription on Purchases

Column (1), (2) and (3) of Table A.6 list the regression results of the three propensity score matching methods we discussed earlier. To account for possible serial correlation, we cluster standard errors at the individual level (Bertrand, Duflo and Mullainathan 2004). As indicated in Table A.6, the effect of subscription on other products' purchase is positive and significant under both static matching and dynamic matching (Column (1) and Column (2)). Both of these two methods show that subscription has a large effects on the purchases of the other products ($\beta = 0.670$, $p < 0.001$ and $\beta = 0.508$, $p < 0.001$). The effect size is smaller but remains significant ($\beta = 0.244$, $p < 0.01$) under look-ahead matching, which controls for unobserved individual time-invariant variables but have to forego the majority of data.[5]

[Insert Table A.6 about here]

As mentioned earlier, the key identifying assumption of our method is that the treatment group and control group have parallel trends in terms of the purchase of not-for-subscription products after matching. Table A.7 presents the results for the placebo tests for the three methods. We do not find evidence showing that the control group and the treatment group differ in their pre-treatment trends as the estimates for the coefficients of the $treated\_fake$ are not significant.

[Insert Table A.7 about here]

The results in Table A.6 show that subscription increases customers' purchase during the subscription period. Thus, subscriptions bring two positive effects for the grocery chain: 1) it locks in customers' future purchases of the subscription products 2) it also has positive spillovers for other products.

For the rest of the analyses, we use the sample from dynamic PSM approach because it can account for the dynamic attributes and also keeps most of the sample data.

Category level effects

Next, we estimate the effects of subscription on purchases of other products in different categories. We estimate Equation 2.4 for each category.

$$\begin{aligned} \log(Category\_spending_{it}) = &\alpha_i + \gamma_t + \beta_1 * SubscriptionSameCategory_{it} \\ &+ \beta_2 * SubscriptionDifferentCategory_{it} + \epsilon_{it} \end{aligned} \tag{2.4}$$

Here, $\log(Category\_spending_{it})$ is the log-transformed spending of not-for-subscription products in one of the 14 categories. $SubscriptionSameCategory_{it}$ is an indicator which equals 1 if Customer i receives subscription from this specific category in Week t. $SubscriptionDifferentCategory_{it}$ is an indicator variable which equals 1 if Customer

[5] We also estimate the causal effect using the causal forest(Wager and Athey 2018) and the results are qualitatively the same.

i receives subscription from other categories in Week t. Other variables are defined similarly as those in Equation 2.3.

Figure A.4 illustrates the regression results. The two dots in each row indicate the effects of subscription inside and outside of the category. For example, the blue dot for "ProcessedMeatAndEgg" indicates that if a customer subscribes chicken eggs, then she purchases less of the other products, say duck eggs, in this category. The red dot for "ProcessedMeatAndEgg" means that if a customer subscribes to some products that are not in this category, say milk, then she purchases more of the other products in "ProcessedMeatAndEgg" category. As subscriptions are offered in only 4 categories, the estimates of the treatment effects from the subscription on other products in the same category are only available for these four categories. However, products in every category could be affected by the subscription of products in other categories. The results in Figure A.4 show a clear split between these two types of treatments: subscriptions of products from the same category will decrease the purchases of other products in the category; while subscriptions in different categories, in most cases, will increase the purchase of products from the focal category. Thus we observe both substitutabilities and complementaries at the category level and the effects roughly depend on whether the subscription is from inside or outside of the category.

[Insert Figure A.4 about here]

2.4.2 Mechanism of How Subscriptions Affect Other Products

In this section, we explore the possible reasons for the increasing purchase of the other products after customers' subscriptions.

There could be multiple different mechanisms how subscription increases customer's purchases of the other products. Due to the data limitation, we are only able to test mechanisms that can be tested by transactions themselves. As the subscription products are delivered regularly to the customers, one critical change due to subscription is the increased number of interactions between customer and the retailer because of the periodic delivery. One natural conjecture is that the delivery of products reminds customers of the retailer. We provide two pieces of evidence to support this conjecture.

Evidence 1: Higher purchase probability in the subscription delivery day. First, if subscription increases customers' purchases of other products due to reminding effects from the product delivery, we should expect that customers are more likely to purchase other products during the day when subscription products are delivered than those days when they are not. To test this, we need to focus on customers' purchases on the daily level. Specifically, we check whether the probability of purchasing is higher in the days when subscription products are delivered, compared to the days in subscription period but without delivery. We estimate the following linear probability model:

$$Purchase_{it} = \alpha_i + \gamma_t + \theta_p + \beta_1 * Subscription_{it} + \beta_2 * Subscription_{it} * Delivery_{it} + \epsilon_{it} \quad (2.5)$$

$Purchase_{it}$ is a binary variable which equals 1 if Customer i purchases the not-for-subscription products on Day t and equals 0 otherwise. $Subscription_{it}$ is the indicator variable which equals 1 if Day t is in the subscription period and 0 otherwise. $Delivery_{it}$ is the indicator for whether there is subscription products delivered in Day t for Customer i. Table A.8 shows regression results. The estimates indicate that the probabilities of purchasing other products are higher in the days when subscription products are delivered ($\beta_2 = 0.601$, $p < 0.001$).

[Insert Table A.8 about here]

Evidence 2: Complementary products increase more in the subscription period. As illustrated in Figure A.4, the effects of subscription are not uniform among the many categories. For grocery, most customers generally derive additional utility by consuming certain products together and these products are complementary to each other. If the subscription delivery reminds customers of grocery needs, such reminding may stimulate the demand more for the products that are complementary to the delivered products than those that are not.

The 23 SKUs with subscription options can be roughly classified into five types: chicken egg, rice, milk, cooking oil, and bottled water. Chicken eggs are usually consumed together with some other products, while the other four types of products can almost be consumed together with any other grocery products. Thus we narrow our focus to subscription for chicken eggs and ask the following question: do the subscriptions to chicken eggs increase the purchase of the products that are complementary to chicken eggs more than those that are not?

To answer this question, we have to first quantify the idea of "complementary" products, which customers usually consume together. In our data set, suppose there are in total N products, then there are $N-1$ relationships between eggs and other products. When N is large, manual coding whether one product is complementary to eggs or not is neither feasible nor reliable. To solve this problem, we use a data mining method called Apriori algorithm (Agrawal and Srikant 1994).

Apriori algorithm has been used extensively to identify association rules between the items in a database. To collect data of customers' grocery consuming patterns, we scrape all the popular family dinner recipes from the largest recipe website (xiachufang.com) in China. We identify ingredients and the popularity (number of times the recipe has been used by others) for each recipe. Figure A.5 shows one example of such a recipe. Then we use the Apriori algorithm to rank the ingredients by their frequencies of appearing together with eggs with the popularity-weighted recipe information. We only consider items that appear frequently and classify those products that frequently coexist with eggs as complementary products with eggs.[6] Figure A.6 illustrates our procedures using Apriori algorithm. Table A.9 lists the most frequent rules.

[Insert Table A.9 about here]

The next step is to map the classification results into our dataset. As ingredients' names and the SKUs' names are not identical, we segment the SKU names into keywords and if there are matches between keywords of an SKU's name and the complementary ingredient's name, we classify the SKU as complementary product to chicken eggs. For example, because green onion is complementary to chicken eggs according to the Apriori algorithm, "XXX green onion 100g package", which contains keywords "green onion", is classified as complementary product to chicken eggs. The other SKUs will be classified as non-complementary to eggs.

Then we estimate the following two econometric models:

$$\log(Spending\_Egg\_Comp_{it}) = \alpha_i + \gamma_t + \beta_1 * Egg\_Subscription_{it} + \epsilon_{itp} \tag{2.6}$$

[6] We use a threshold that support is at least 2%, and confidence is at least 20% for the association rules that can be classified as complementary. Roughly speaking, The support of an itemset X is the probability of records containing X, which describes the the popularity of itemset X. The confidence is the probability of records containing both X and Y, conditional on that itemset X is present.

, and

$$\log(Spending\_Egg\_NonComp_{it}) = \alpha_i + \gamma_t + \beta_1 * Egg\_Subscription_{it} + \epsilon_{itp} \tag{2.7}$$

, where $Spending\_Egg\_Comp_{it}$ and $Spending\_Egg\_NonComp_{it}$ are the overall spending on all the SKUs that are complementary to eggs and non-complementary to eggs respectively. $Egg\_Subscription_{it}$ is the indicator which equals 1 if Customer i in Week t subscribe to chicken eggs. Other variables are defined similarly as those in Equation 2.3.

[Insert Table A.10 about here]

Table A.10 lists the regression results. First, the coefficients in the two models ($\beta_1 = 0.514$, $p < 0.001$ and $\beta_1 = 0.430$, $p < 0.001$) are both positive, which means subscriptions of chicken eggs have positive spillovers on all the other products. Second, the effect of chicken eggs' subscription on the complementary products of chicken eggs is larger than that of the non-complementary products (Chow-statistic: $1.049 > 0.105$, the critical value). The larger increase of complementary products suggests that customers purchase more other products during subscription partly due to the needs to complement chicken eggs. This supports that subscription increase customers' purchase of other products because of reminding effect.

2.5 Conclusions and Discussions

Given the rising popularity of product subscription in the business world, the literature seems to fall behind. The consequences of product subscription are still not well-understood and the previous researches are too limited to provide enough insights and guidance for practitioners. Product subscription requires new operations and customer management arrangements, and may interact with the company's original marketing decisions. Thus, companies are also faced with huge challenges, despite the high potential of product subscription. Experimentation without the necessary insights about product subscriptions would put the companies in a risky position. Our paper seeks to contribute knowledge to this new business phenomenon of product subscriptions.

By cooperating with a membership-based grocery chain which experiments with product subscriptions, we investigate how product subscriptions affect customers' purchases of the other products. The granular individual-level data allow us to identify the causal impact of subscription on individual purchases. In summary, we find that subscriptions increase customers' overall purchases of the other not-for-subscription products. The overall complementarity between subscription products and other products can be decomposed into two opposing effects: the substitution effect on the purchases of the products in the same category and the complementary effects on the purchases of the products in the other categories. We also show that the probability for customers to make a purchase for other products is higher on the day of subscription delivery, and purchases increase more for the products that are complements to the subscribed product. The evidence suggests that subscriptions increase customers' purchases of other products by reminding customers of the retailer and their grocery needs through subscription delivery. Given that the complementarity between products is generally present in many different industries, we believe that our results can be applied to other contexts besides grocery.

To the best of our knowledge, our paper is the first empirical research to study product subscriptions. Our research contributes on multiple fronts. First, we show that the subscription of products increases the purchases of other products. The results assure some practitioners who fear the negative effects of subscriptions on the rest of the products. Second, the category-level effects we find provide insights for implementing subscription in a multi-product context. The heterogeneous effects of subscriptions provide businesses with some initial guidance for strategic design of subscription plans, e.g. which products to pick for subscription to maximize overall revenue. Given the negative effects on products within the same category, one intuitive implication is that subscription may not be well suited for retailers whose assortments are not well diversified to provide enough opportunities for complementarity. Third, the mechanism we discover provides a deeper understanding of product complementarities for a multi-product firm. We show evidence to support that product subscriptions complement other products by reminding customers through subscription delivery. In this case, the products themselves serve as informational channels for the firm. The subscriptions provide additional opportunity for interactions between customers and retailer, and increase customers' purchases. The reminding effect of subscription also infers that retailers can benefit from subscriptions that are frequently delivered and that the delivery cycle are aligned with the cycle of customers' purchase decisions.

The complementarity also requires that store managers think of the marketing actions for different products and different categories holistically. Here in our context, given that the subscription increases purchases of other products and products in other categories, the retailer could motivate more customers into subscription by promotions, such as prices discount for the subscription or more advertisements to direct customers' attention to subscriptions. At the same time, the compensation plans designed for salespeople should also be based on a broader range of indicators. If a salesperson is only rewarded by the sales of the subscription, the salesperson may be under-incentivized given the positive spillovers of subscription. A better-designed compensation plan should align the motivation better with the company's overall interest.

The key limitation of our research comes from the lack of competition information. Because we have no data on customers grocery purchase in other retailers, we are not able to say exactly whether the increased purchases are because of stealing from the competitors or the general expansion of overall grocery needs. In a highly competitive business world, a store manager may have to worry more about the competition from the other retailers than the competition among different categories and products her stores offer. Thus, even if subscription reduces purchases of other products in the same category, some single-category firm may still roll out subscription plan to compete with other firms. However, without the information of customers' purchases from other retailers, we are not able to directly contrast the effects on the cross-retailer competition with the effects on cross-product/cross-category competition.

Subscription could also have more subtle impacts on customers' purchases beyond quantity change. One possible change due to subscription is customer's exploratory behavior. As subscriptions fix part of customer consumptions for a relatively long period of time, will subscription also reduce the diversity of the other products that the customer purchases? Or instead, would customers seek more varieties of the other products to compensate for the subscription. These subtle changes could have profound consequences for not only marketers but also regulators, as exploration behaviors have been shown to be related to de facto entry barriers (Schmalensee 1974).

We only show the evidence for subscriptions to function as product fulfillment channel and information channel. However, subscription could also serve as a promotional channel, through which customers will also purchase more of the other products. The new option of subscription may suit a customer's need more than the traditional buying option, resulting in a higher satisfaction level towards the retailer and perhaps more purchases of the not-for-subscription products. As a form of a long-term contract, subscription may establish a long-term relationship between the customers and the retailer, increasing customers' preference for the retailer and reducing customer churn. Future researches could explore this promotional channel for product subscriptions.

Chapter 3

Essay II: The Effects of New Subscription Plans on The Product

3.1 Introduction

In Chapter 2, we examine how subscription affects customers' purchases of other products. Although the products that are available for subscription account for only a small percentage of all the products in our context, knowing how they are affected by being made available for subscription helps us build a holistic view of the effects of subscription. The understanding of how subscriptions affect the focal products also helps to generalize our findings to the context where most, or even all products, are for subscription. Actually, for many of the prominent examples of the product subscription business, subscription is the main or only option.

We find that subscribing to products actually reduces the purchases of other products in the same category in Chapter 2. Given that products in the same category are usually similar in terms of the demand they fulfill, we should expect a similar effect for subscription on the rest of the purchases (excluding subscription) of the focal product. While a more interesting question is: will introduce subscription for a specific product increase the purchase of the product? Actually, this might be the first question marketers may ask when they are considering launching subscription delivery plan for some of their products. The primary goal for many retailers, like Amazon, to introduce subscription plan for a certain product, e.g. pistachio, is to boost the sales of the product. Unfortunately, there is not yet an answer to this empirical question despite that a large amount of businesses are considering jumping on the bandwagon.

In this paper, we investigate this question using the same dataset described in Chapter 2. To be specific, we study whether customers' purchase of a product would change had the retailer not introduced subscription plans for the product. We compare customers' purchases of the products in two business scenarios:

1. Hybrid: one-time purchase + subscription.

2. One-time Purchase Only.

We suppress the option of "subscription only" in this paper as this option, although emerged as the only purchase option for many retailers, is unlikely to be available in our context where the grocery chain is competing for the mass market. We focus on whether a retailer could introduce subscription plans, i.e. use Hybrid, for some of their products, and benefit the sales of the products.

Although we have data from both the period when the retailer is in the second scenario and data from the period when the retailer is in the first scenario, we are unable to draw conclusions by simply comparing customers' behaviors in these two periods due to the lack of control group. To have a better "control", we build a structural model that takes consumers' dynamic choices into consideration. Next, we estimate the model using the data from the period when customers are in the Hybrid scenario and simulate a counterfactual scenario where customers have only the option of a one-time purchase. Then we have a clean comparison where everything, except the available purchase options, is held unchanged.

Our results show that the introduction of subscription plans does boost the overall purchase of the product. Both large families and small families purchase more eggs in the Hybrid scenario than in the One-time Purchase Only scenario. However, the effects of subscription plans on different options of one-time purchases are not uniform. Overall, we find that customers' purchase of chicken eggs will halve (49.9%) had the retailer not introduced the subscription plans for chicken eggs. The ratio of dollar sales is similar (50.4%). These results show the significant effects of subscription plans on the sales of the product.

3.2 Model and Data

In our context, subscription is a form of investment: customers subscribe to products by paying a lump sum price upfront and recover the cost through future consumptions. Thus, customers' choices of subscription should have taken both the current and the future situations into consideration. Their goals of subscribing is also to maximize their utility in a relatively long period of time, instead of just the moment of purchase. Thus, customers' choices are collectively determined by the states in the specific period and the choices made previously. In the grocery context, customers' decisions are usually made by themselves/within a household and interactions between customers should not be prominent. Given these particular features of the subscription context, we build our model based on the classic single agent dynamic programming framework (Rust 1987, 1994) .

Our model is similar to the stockpiling models proposed by Hendel and Nevo (2006) and Erdem, Imai and Keane (2003) given the similar nature of stockpiling behavior and subscription. Product subscription can be viewed as customers "stockpiling" the bulk of the products in the retailer's warehouse after purchasing them. The key difference lies in that subscription avoids the inventory cost of holding physical commodities. Moreover, consumers' consumptions are bound by the specific delivery plan in the subscription. In comparison, consumers are only limited by the overall inventory levels in the stockpiling context. Stockpiling usually happens in the domain of durable products; while product subscriptions are most widely seen in the domain of perishable products (or products that require frequent and regular replacement). Product subscription also enables customers to consume products that are fresh and new.

Although we find complementarity between subscription products and other products in Chapter 2, we decide not to model the relationship as part of consumers' utility function. For subscription products, consumers themselves may not be aware of their future needs for products that are complementary to the subscription products. It is reasonable to assume that customers subscribe due not to the consideration that subscription delivery in the future will remind them of the retailer and the products to consume together with the subscription products. We do admit that different products consumers purchase in the same basket are related to each other. However, a complete model that incorporates all products' relationships is almost impossible to be fully estimated, given the huge number of possible combinations of different products. Thus, we compromise by focusing on the purchase, subscribing , and consumption of a single product (here we choose chicken eggs) and assume customers make decisions by optimizing locally for this specific product.

We model consumers' decision of subscriptions as the outcomes of prices changes of the subscription options and the one-time purchases, and the transaction cost associated with each purchase, e.g. the cost to go to an offline store/website to place the order. As we discussed in Chapter 2, both lower unit prices[1] and transaction cost could be the reasons why customers choose subscription over one-time purchase. In our context, transaction costs include both the tangible transaction costs, i.e. transportation cost, handling cost, and the intangible transaction costs, i.e. traveling time, cognitive resources used to make decisions, that occure for each shopping trip. In each period, the consumer's flow utility consists of (1) the consumption utility, (2) transaction cost, (3) purchase utility. We simplify consumer's consumption by assuming that consumer consumes all products she purchases and those subscription products delivered. This simplify our model by ignoring the effects of inventory has on customers' purchase decisions. We believe this assumption is suitable in the grocery context where products are usually perishable[2] .

We borrow the techniques and assumptions mainly from Hendel and Nevo (2006) as their model is more general and required less restrictive assumptions compared with other approaches. At the beginning of each period, customers decide which option to choose based on the current states, in order to maximize their long-term discounted utility flow.

[1] For chicken eggs, the unit prices of eggs in subscription range from 90.7% to 96.4% of prices of one-time purchases of the same type of eggs.

[2] For the product we choose to estimate our model, chicken egg, the shelf life is roughly two weeks.

The consumer's problem is formulated as:

$$V(l_1, d_1) = \max_{\{k_t\}_{t=1,2,...}} \sum_{1}^{\infty} \beta^t E[u(c_{ht}; \theta_h) - tc_{k_t} + \underbrace{\alpha_h p_{k_t t} + \xi_{hk_t t} + \epsilon_{hk_t t}}_{\text{purchase utility}} | l_1, d_1]$$

$$\text{s.t.} \quad c_t = c(d_t, q_t) = q_t + \sum_i d_{it}$$

$$l_{it} = l(l_{i,t-1}, k_{t-1}) = \begin{cases} l_{i,t-1} + w_{k_{t-1}} - 1 & \text{if} \quad k_{t-1} = i, \\ \max\{0, l_{i,t-1} - 1\} & \text{if} \quad k_{t-1} \neq i \end{cases}$$

$$d_{it} = d(d_{i,t-1}, l_{i,t-1}, k_{t-1}) = \begin{cases} q_{k_{t-1}} & \text{if} \quad k_{t-1} = i, \\ d_{i,t-1} & \text{if} \quad k_{t-1} \neq i \& l_{i,t-1} > 1, \\ 0 & \text{if} \quad k_{t-1} \neq i \& l_{i,t-1} \leq 1 \end{cases} \tag{3.1}$$

, where l_{it}, d_{it} stand for the number of periods left of subscription delivery[3] and quantity of delivery in the beginning of week t of subscription product i respectively. We choose a week as a unit period as the period of chicken eggs subscription is a week. $l_1 = \{l_{i1}\}_i$, and $d_1 = \{d_{i1}\}_i$ are vectors of all l_{it} and d_{it} at the beginning of Week 1. β[4] is the discount factor. Consumer h chooses the SKU k_t in the beginning of each period to maximize her overall discounted utility. Each SKU k corresponds to a 3-tuple (j_k, q_k, w_k) with brand j ,its specific quantity q and periods of delivery w. For example, if the customer makes no purchase in a period, then $k = 0$ and $q = 0, w = 0$ accordingly . $w > 1$ if the SKU is subscription, and $w = 1$ if the SKU is one-time purchase. c_{ht} is the total quantity of a product consumed (subscription delivery and one-time purchase) of customer h in period t and u is the consumption utility function. tc_{k_t} denotes the transaction cost which is equals zero if $k = 0$ and is positive otherwise. $p_{k_t t}$ is the price of SKU k_t in period t and $\xi_{hk_t t}$ is SKU-specific taste that could be related to brand, size, customer and could vary across time (Hendel and Nevo 2006). ϵ is the SKU and time specific idiosyncratic shock that is i.i.d. type I extreme value (Rust 1987; Berry, Levinsohn and Pakes 1995; Hendel and Nevo 2006).

Note that we assume that consumers purchase only one brand of product in a single

[3] An alternative approach is to model subscription exactly as stockpile as quantity "stockpiled" in the retailer's warehouse. We choose the current approach due to its less number of states.

[4] I use $\beta = 0.9$ for my estimation.

week in our model, which is not always the case in the data[5] . As the proportion of such cases is very low, we modify these cases to suit our model[6] . To reduce the number of states, we eliminate the weekly sizes that appear less than 10 times and modify the size as major size with the least difference, and the paid prices are adjusted accordingly[7] .

According to this model, products differentiation only occurs during the moment of product purchase as reflected by the structure of the purchase utility in 3.1. After purchase, products of different brands are homogeneous and this is why the consumption utility u is only dependent on the total quantity consumed. This model fits the chicken eggs context: although the eggs of different brands and sizes may differentiate at the moment of purchase due to the different packaging and presentations, they can hardly be distinguished after being cooked.

Consumers' decisions are dynamically related through the subscription statues, i.e. l_t and d_t. For example, if a consumer makes an egg-subscription that delivers 12 eggs for 12 weeks, then the customers will receive 12 eggs per week in the week and 11 weeks that follow. Thus, at the beginning of the next 11 weeks, the customer has to take into consideration that 12 eggs will be delivered (and consumed) even if she does not make any purchases of eggs. This idea is translated to the period-to-period state transition relationship described in the customer's problem 3.1. To simplify the model, we treat the case of subscribing to a product while the last subscription of the same product has not ended as simply extending the subscription period. This allows us to track only one (l_i, d_i) state for each subscription product i.

As mentioned earlier, we choose chicken eggs as the product to estimate our model. Choosing only the transaction data for chicken eggs is a compromise trading off the tractability/complexity of the model and the credibility of our estimation results. There are several unique features of chicken eggs that are well suited for our structural model.

[5] There are purchases of more than one brand of eggs in only 1.2% of the consumer-week cases.

[6] We modify the orders as follows: if there are one-time purchases of different brands, we use the brand that accounts for the most quantity and aggregate all the quantity; if there are subscription and one-time purchases, we delete the one-time purchases; if there are two subscriptions, we aggregate them as one single subscription of the brand that accounts for the most quantity, the quantities are aggregated and the lengths of delivery are averaged.

[7] For example, the subscription option "12 eggs, 24 weeks" are only chosen for eight times. For each incidence, we treat that purchase as the subscription option "12 eggs, 12 weeks" and halve the prices accordingly.

First, chicken eggs are generally popular[8] and the demand is relatively stable as consumers will not easily get satiated with them. Thus, they are likely to be purchased regularly and periodically. second, chicken eggs cannot be stored for a long time of period and have to be consumed in a relatively short period of time, which suits our structural model with simplified consumption choices assumptions. Third, chicken eggs are less differentiated among different brands in the consumption stage, compared with other products, say detergent. As a results, the model assumption that brands only differentiate in the purchase stage is more realistic.

To ensure that different purchase options are available through the whole estimation period, we choose the transactions of eggs brands that offer subscription options and sell in each of the 94-week period. There are in total 12 SKUs in 2 brands that satisfy this creteria.

Weekly price p is calculated as the average price in a certain week. We assume that weekly prices change exogenously according to a first order Markov process and this process is known by the customers. As can be seen in Figure A.7, the time series of prices of different chicken egg SKUs are relatively stable without obvious drift or seasonality. It is also likely to be the case that customers have only limited memory about the history of chicken eggs as they do not cost a substantial amount of money. Thus, the first order Markov process, which assumes that customers only remember the prices in the last week, seems acceptable as the model for price variations in our case.

[Insert Figure A.7 about here]

Using Bellman equation, the consumer's problem 3.1 can be represented as:

$$V(l_t, d_t) = \max_{\{k_t\}} \left\{ u(c_{ht}; \theta_h) - tc_{k_t} + \alpha_h p_{k_t t} + \xi_{hk_t t} + \epsilon_{hk_t t} + \beta E[V(s_{t+1})|l_t, d_t, k_t] \right\} \quad (3.2)$$

3.3 Estimation

The estimation is based on the framework proposed by Rust (1987). The basic idea is to search for the parameters that maximize the likelihood of observed customers' choices,

[8] Chicken eggs account for more than 95% of the amount of all eggs sold in our dataset.

assuming that the customers are solving a dynamic programming problem. In our case, the likelihood we want to maximize is:

$$Pr(k_1, k_2, ..., k_T | P_1, P_2, ..., P_T; \theta)$$

Here, P_t represents the price vector of all SKUs in period t.

Our estimation follows the three-step approach in Hendel and Nevo (2006) closely. The approach simplifies the multi-period product choice model into two separate stages: in the first stage, customers decide what size to purchase; in the second stage, customers choose the brand to purchase conditional on the size decided in the first stage. This simplification relies primarily on the assumption that product differentiation occurs only in the purchase phase.

Mathematically, if we decompose each SKU k_t into brand j and size x (each x corresponds to a combination of quantity q and weeks of delivery w), the SKU choice can be viewed as a composite of brand choice and size choice:

$$Pr(j_t, x_t | P_t, l_t, d_t) = Pr(j_t | x_t, P_t, l_t, d_t) \times Pr(x_t | P_t, l_t, d_t)$$

The two components can be computed individually because (using Equation 3.2):

$$\begin{aligned} Pr(j_t, x_t | P_t, l_t, d_t) &= \frac{\exp(\alpha p_{jxt} + \xi_{jx} + u(x_t, d_t) - tc_{x_t} + \beta E(V(l_{t+1}, d_{t+1}) | x_t, P_t, l_t, d_t)}{\sum_{j', x'} \exp(\alpha p_{j'x't} + \xi_{j'x'} + u(x', d_t) - tc_{x'} + \beta E(V(l_{t+1}, d_{t+1}) | x', P_t, l_t, d_t)} \\ &= \frac{\exp(\alpha p_{jxt} + \xi_{jx} + M(j, x, P_t, l_t, d_t))}{\sum_{j', x'} \exp(\alpha p_{j'x't} + \xi_{j'x'} + M(j', x', P_t, l_t, d_t))} \end{aligned}$$

, where $M(j, x, P_t, l_t, d_t) = u(x_t, d_t) - tc_{x_t} + \beta E(V(l_{t+1}, d_{t+1}) | x_t, P_t, l_t, d_t)$. Because M is independent of brand choice j, $M(j, x, P_t, l_t, d_t) = M(x, P_t, l_t, d_t)$. We simplify the brand choice as:

$$Pr(j_t | x_t, l_t, d_t, P_t) = \frac{\exp(\alpha p_{jxt} + \xi_{jx})}{\sum_{j'} \exp(\alpha p_{j'xt} + \xi_{j'x})} \tag{3.3}$$

After the simplification, the conditional brand choice model is a static brand choice using only the brands offered in the size that is purchased by the customer.

Another hurdle in the estimation is the large number of price states (one price state for each SKU). In the second step, we reduce the number of prices states by aggregate SKU-specific price information to size-specific price. Using the estimates from Step 1, we calculate the size-specific inclusive values as:

$$\rho_{xt} = \log\{\sum_{j'} \exp(\alpha p_{j'xt} + \xi_{j'x})\} \tag{3.4}$$

The inclusive value can be viewed as a weighted price index of a specific size. It is the "price" that customers can expect by choosing to purchase a specific size x. This is because:

$$E_{\epsilon|x,l,d}\{\max_j[\alpha_h p_{jxt} + \xi_{jxt} + \epsilon_{hjxt}]\} = \log[\sum_j \exp(\alpha_h p_{jxt} + \xi_{jxt})] \tag{3.5}$$

, given the distribution assumption for ϵ.

If we further assume that $F(\rho_t|P_{t-1})$ can be summarized by $F(\rho_t|\rho_{t-1})$ (Assumption 4 in Hendel and Nevo (2006)), where ρ_t is the vector of inclusive values of all sizes, then we need to only track the transition process of ρ_t. This reduces the number of price states from the total number of SKUs to the number of sizes.

In Step 3, we solve the simplified dynamic programming problem in which customers decide on the size:

$$V(l_t, d_t, \rho_t, \epsilon_t; \theta) = \max_{\{x\}}\{u(x,d) - tc_{xt} + \rho_{xt} + \epsilon_{xt} + \beta E[V(l_{t+1}, d_{t+1}, \rho_{t+1}, \epsilon_{t+1}|l_t, d_t, \rho_t, \epsilon_t, x)]\} \tag{3.6}$$

Please refer to the Appendix of Hendel and Nevo (2006) for proof for the equivalence of the likelihood computed from the original problem and this simplified problem.

The estimation of this simplified problem follows the nested algorithm proposed by Rust (1987). The basic idea of the nested algorithm is that the outer loop searches for a set of parameters to maximizing the likelihood of the observations; In the inner loop, for a given set of parameters, calculate the expected value function by contraction mapping and compute the conditional choice probabilities using the expected value functions.

According to Equation 3.6, define function EV as:

$$EV(l_t, d_t, \rho_t, x_t; \theta) \equiv E_{\rho_{t+1},\epsilon}[V(l_{t+1}, d_{t+1}, \rho_{t+1}, \epsilon_{t+1}; \theta)|l_t, d_t, \rho_t, x_t]$$

If we assume that the price process ρ is independent of the shock ϵ (conditional independence), we have:

$$\begin{aligned}
EV(l_t, d_t, \rho_t, x_t; \theta) &= E_{\rho_{t+1},\epsilon|l_t,d_t,\rho_t,x_t}\{\max_{x_{t+1}}[u(x_{t+1}, d_{t+1}; \theta) - tc_{x_{t+1},t+1} + \rho_{x_{t+1},t+1} + \epsilon \\
&\qquad + \beta EV(l_{t+1}, d_{t+1}, \rho_{t+1}, x_{t+1})]\} \\
&= E_{\rho_{t+1}|l_t,d_t,\rho_t,x_t} E_{\epsilon|\rho_{t+1},l_t,d_t,\rho_t,x_t}\{\max_{x_{t+1}}[u(x_{t+1}, d_{t+1}; \theta) - tc_{x_{t+1},t+1} \\
&\qquad + \rho_{x_{t+1},t+1} + \epsilon + \beta EV(l_{t+1}, d_{t+1}, \rho_{t+1}, x_{t+1})]\} \\
&= E_{\rho_{t+1}|\rho_t} \log[\sum_{x_{t+1}} \exp\{u(x_{t+1}, d_{t+1}; \theta) - tc_{x_{t+1},t+1} + \rho_{x_{t+1},t+1} \\
&\qquad + \beta EV(l_{t+1}, d_{t+1}, \rho_{t+1}, x_{t+1})\}] \\
&= \int_{\rho_{t+1}} \log[\sum_{x_{t+1}} \exp\{u(x_{t+1}, d_{t+1}; \theta) - tc_{x_{t+1},t+1} + \rho_{x_{t+1},t+1} \\
&\qquad + \beta EV(l_{t+1}, d_{t+1}, \rho_{t+1}, x_{t+1})\}] p(\rho_{t+1}|\rho_t)
\end{aligned}$$

The next-to-last equality uses the property of type I extreme value. For $\beta < 1$, there is a unique solution for EV_θ. Then we can compute the conditional choice probability as:

$$P(x|\rho, l, d) = \frac{\exp\{u(x, d) + \rho_x + \delta EV_\theta(\rho, l, d, x)\}}{\sum_{x'} \exp\{u(x', d) + \rho_{x'} + \delta EV_\theta(\rho, l, d, x')\}} \tag{3.7}$$

Thus, for each set of parameters θ, we can obtain the overall probability of observing the choices customers make given the states. In the outer loop, we search over parameters space to maximize this probability.

3.3.1 Step 1: brand choice

We use only the data from customers who subscribed at least once as those who never subscribe may not be suitable for a dynamic model that assumes agents are forward-thinking. The data contains customers' shopping history, including both one-time purchase and subscription, of chicken eggs during the weeks when the chicken eggs subscriptions are available.

[Insert Table A.11 about here]

Table A.11 presents the results of conditional logit regression of brand choices conditional on the size. We test three different specifications. Column (1) presents the simple regression results. Column (2) presents the results for regression that incorporates customers' heterogeneity by interact price with prominent customer's attributes. Column (3) includes brand heterogeneity by adding brand fixed effects. The results show that male customers (gender = 1) are less price-sensitive than female customers. Whether a customer is relatively new or whether the family size is larger than three are not significant. We use the regression results in Column (3) to calculate our Step 2 inclusive values.

3.3.2 Step 2: inclusive values

The second step calculates the size-specific inclusive values for each week using Equation 3.4. Because all the inclusive values are continuous, we also need to discretize them as the input for the third step. In our context, one size could consist of more than 20 times the number of eggs than the other size, which means the inclusive values could also differ by more than 20 times. If we discretize the inclusive values using the same scale, many states will end up useless[9] . Thus, we discretize each inclusive value based on the values it could take. Then we calculate the transition probability using the price history[10] . A similar method has been used in Rust (1987).

3.3.3 Step 3: size choice

To estimate the simplified dynamic programming problem, we have to choose certain functional forms for the model. Our focus is the utility consumer derives by consuming a certain amount of eggs, i.e. u. We assume that $u(c) = \alpha_1 c + \alpha_2 c^2$. The quadratic function allows for concavity/convexity of utility. Transaction cost tc is a constant if customers make a purchase in a certain week and are zero otherwise. We follow the value iteration method proposed in Rust (1987). Table A.12 presents the estimated parameters. The results indicate that larger families derived lower utility from the same amount of eggs. This could be due to the fact that the same amount of eggs will

[9] For For example, if we discretize all the inclusive values using a step of 0.01, 0 being the first state, then subscription size of 12 eggs/week, 12 weeks will not be in the first 1000 states as its price has never fallen below 100 RMB.

[10] Alternative methods include kernel estimation (e.g. using a Gaussian kernel to estimate the transition probability), and estimate a linear model using the lagged states of all sizes (Hendel and Nevo 2006).

be divided among more people. Thus, large families need to purchase more eggs to achieve the same level of utility compared with small families.

[Insert Table A.12 about here]

3.4 The Effects of Product Subscription on Purchase

In this section, we perform counterfactual analysis to figure out the effect of the introduction of a subscription plan on the sales of the product. We use the historical prices of eggs and our estimated structural model from Section 3.3 to simulate customers' purchase choices over the same period of time. Table A.13 presents a summary of the total number of choices of different sizes in the original scenarios (Columns label as "Hybrid") and in the simulated scenarios without product subscription options (Columns labeled as 'One-time Only").

[Insert Table A.13 about here]

The comparisons show that the total number of eggs purchased by customers of both small families and large families will decrease if the subscription options were to be eliminated. The decrease does not come from more incidences when customers make no purchase as the number of times choosing "0 eggs" remains more or less the same. Instead, the decreases come from the shift from subscription plans to different one-time purchases. Interestingly, the simulated results indicate that eliminating the subscription options has distinct effects on different sizes. Elimination of subscription options increases the purchases of large sizes ("18 eggs" and "24 eggs"), while decreasing the purchases of small sizes ("6 eggs" and "12 eggs"). This could be due to relationships between different sizes: 'different is a close alternative to "24 eggs, 12 weeks" and "12 eggs, 12 weeks"+"12 eggs"; and "18 eggs" is a close alternative to "12 eggs, 12 weeks" + "6 eggs". Overall, the total number of eggs is only 49.9% of the current case should there be no subscription options for chicken eggs. The total dollar sales are only 50.4% in the scenarios without subscription options.

However, our result that subscription plans doubles the sales of the focal products has to be interpreted with caution. One key limitation of our approach is the data we used, i.e. only transactions of chicken eggs from the two popular brands. The results does not translate to that customers increase their egg consumption, and thus the protein intake, by a huge amount because customers may simply shift their purchases from other retailers. neither do our results indicate that subscription plan increase the overall sales of the whole category. The large increase in the purchase of chicken eggs may be at the expense of the decreases of other egg products or other chicken egg brands. Customers may simply switch their purchases of egg products to chicken eggs and the overall demand for eggs and egg-related products may not necessarily increase. Actually the decrease of the sales of other produts in the same category is shown in Chapter 2. Some well-designed field experiments may be needed to fully understand the broader effects of subscription plans.

3.5 Conclusions and Discussions

In this paper, we investigate how subscription plans affect the purchase of a product. This question is important for many marketers to consider the subscription plan in the first place. We build a structural model which incorporates the dynamic nature of subscription plan and customers' considerations among the options of subscription and one-time purchase. The simulated counterfactual scenario without subscription plans provides us a great control to examine the effects of subscription plans. Our results reveal the remarkable efficacy of the subscription plan: the total number of eggs purchased in the scenario with the subscription is more than double compared with the scenario where only one-time purchases are available. These positive results depict the promising outcomes that subscription plans could bring and the boost of the sales of the product itself might be the reason why product subscription is becoming more and more popular.

However, we do find that subscription options affect various one-time purchase options differently. As subscription options become available, the one-time purchase options with smaller sizes are positively affected while those with larger sizes are negatively affected. The uneven influences of subscription plans mean that retailers cannot simply introduce subscription plans blindly. In a world where large sizes are initially dominant, subscription plans could negatively affect the sales of the large sizes and possibly reduce the overall sales/revenue.

Retailers should also be cautious before launching subscription options as they complicate the process for retailers to make decisions. The existence of subscription plans increases the linkage between business decisions between different periods. The retailers have to now put their assortment decisions, i.e. what to sell, and the pricing decisions, i.e. how much to sell for, in a dynamic framework. Any marketing decisions in one period could have long-lasting effects that are hard to undo. This increases the cost of marketing experimentations and making mistakes. Thus, although subscription plans that are well implemented will likely increase the profitability of the business, they pose a greater challenge to the management ability of firms. This could be the reason why most retailers, except those that are tech-savvy, are still hesitant to apply the subscription model to their businesses.

Chapter 4

Conclusions and Discussions

In my dissertation, I study the effects of product subscription on both the other products and focal product. Given the increasing degree of presence of this new business model and its high potential, researches concerning this specific business model are still lagging behind. By cooperating with a grocery chain which makes some of its products available for subscription in the past few years, I have the opportunity to study this new business model directly using empirical data. I find that product subscriptions have positive effects on both the other products and the focal product. However, the effects are not homogeneous in both cases. For the impact on the purchases of the other products, I show that subscriptions have positive effects for other products cross-category while they have negative effecs for other products within category. For the focal product, subscription show opposite directions of impact on different sizes. I also present evidence that product subscriptions could function as an information channel for the retailer.

Although our results demonstrate the lucrative potential of product subscriptions, they also show the great challenges in implementing a successful product subscription plan to realize the potential. In a business environment where retailers have to manage the products, pricing, promotion, and delivery of thousands of different SKUs and deal with a large number of customers, any marketing actions on a single SKU could result in long-term effects on all other products. Thus, retailers have to take product subscription as part of their overall business strategy and coordinate product subscription with other business decisions of the firms. Instead of just using product subscription as a method to boost the sales of a product, company ought to treat it as an element of the package offered to customers and have the mind set of global optimization. As the empirical evidence indicates, product subscription have the information-delivery effects similar to billboards and online advertisement. Thus, it is acceptable even if subscriptions do not increase sales of the product or category. The temporary loss may be treated as investment.

Product subscriptions pose great challenges for retailers. The design of subscriptions, including the pricing, the interval of delivery, what products to make available for subscription, how to promote the subscriptions plans, have to take into considerations both the products for subscription and other products. Given the high complexity that subscription plans involve, more insights and better understandings are needed in the future for subscriptions to become an effective tool for marketers.

More quantitative researches are still needed for product subscriptions. Given the complicated structure (length, quantity, prices and etc.) of product subscription plans, more researches are required to look into the optimal subscription structure. Researches could study the design of a price menu containing both subscriptions and one-time purchases to maximize retailers' interests. Future researches could also examine how product subscriptions affect the competitive environment. Subscription plans may change both customers' choices of products and retailers. As subscriptions lock in customers for considerable amount of time, they could have profound and long-lasting effects on the competition of the retailers.

References

Agrawal, Rakesh and Ramakrishnan Srikant (1994), "A fast algorithm for mining association rules," *Proceedings of the 20th VLDB Conference*, 513–516.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks.," *Proceedings of the National Academy of Sciences of the United States of America*, 106 (51), 21544–21549.

Avery, Jill, Thomas J Steenburgh, Johnv Deighton, and Mary Caravella (2012), "Adding Bricks to Clicks: Predicting the Patterns of Cross-Channel Elasticities Over Time," *Journal of Marketing*, 76 (3), 96–111.

Bapna, Ravi, Akhmed Umyarov, and Jui Ramaprasad (2018), "Monetizing freemium communities: Does paying for premium increase social engagement?," *MIS Ouarterly*, 42 (3), 719–735.

Barro, Robert J and Paul M Romer (1987), "Ski-Lift Pricing , with Applications to Labor and Other Markets," *American Economic Review*, 77 (5), 875–890.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–890.

Bertrand, M., E. Duflo, and S. Mullainathan (2004), "How Much Should We Trust Differences-In-Differences Estimates?," *The Quarterly Journal of Economics*, 119 (1), 249–275.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Duncan Simester (2011), "Goodbye Pareto

Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," *Management Science*, 57 (8), 1373–1386.

Cachon, Gérard P. and Pnina Feldman (2011), "Pricing services subject to congestion: Charge per-use fees or sell subscriptions?," *Manufacturing and Service Operations Management*, 13 (2), 244–260.

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003), "Why don't prices rise during periods of peak demand? Evidence from scanner data," *American Economic Review*, 93 (1), 15–37.

Chintagunta, Pradeep K (1993), "Investigating Purchase Incidence , Brand Choice and Purchase Quantity Decisions of Households Author (s): Pradeep K . Chintagunta INVESTIGATING PURCHASE INCIDENCE , BRAND CHOICE AND PURCHASE QUANTITY DECISIONS OF HOUSEHOLDS," *Marketing Science*, 12 (2), 184–208.

Danaher, Peter J. (2002), "Optimal Pricing of New Subscription Services: Analysis of a Market Experiment," *Marketing Science*, 21 (2), 119–138.

Datta, Hannes, George Knox, and Bart J. Bronnenberg (2018), "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery," *Marketing Science*, 37 (1), 5–21.

DellaVigna, Stefano and Ulrike Malmendier (2006), "Paying not to go to the gym," *American Economic Review*, 96 (3), 694–719.

Erdem, Tülin, Susumu Imai, and Michael P. Keane (2003), "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1 (1), 5–64.

Fan, Ming, Subodha Kumar, and Andrew B. Whinston (2009), "Short-term and long-term competition between providers of shrink-wrap software and software as a service," *European Journal of Operational Research.*

Guo, Zhiling and Dan Ma (2018), "A model of competition between perpetual software and software as a service," *MIS Quarterly: Management Information Systems*, 42 (1), 101–120.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342.

Hendel, Igal and Aviv Nevo (2006), "Measuring the Implications of Sales and Consumer Stockpiling Behavior," *Econometrica*, 74 (6), 1637–1673.

Kumar, Naveen, Liangfei Qiu, and Subodha Kumar (2018), "Exit, Voice, and Response on Digital Platforms: An Empirical Investigation of Online Management Response Strategies," *Information Systems Research*, 29 (4), 849–870.

Lambrecht, Anja and Bernd Skiera (2006), "Paying Too Much and Being Happy About It," *Journal of Marketing Research*, 43 (2), 212–223.

Liu, Hongju, Pradeep K. Chintagunta, and Ting Zhu (2010), "Complementarities and the Demand for Home Broadband Internet Services," *Marketing Science*, 29 (4), 701–720.

Ma, Dan and Abraham Seidmann (2015), "Analyzing software as a service with per-transaction charges," *Information Systems Research*, 26 (2), 360–378.

McKinsey (2018) "Thinking inside the subscription box: New research on e-commerce consumers."

Mitchell, Bridger M. (1978), "Optimal pricing of local telephone service," *American Economic Review*, 68 (4), 517–537.

Nunes, Joseph C. (2000), "A Cognitive Model of People's Usage Estimations," *Journal of Marketing Research*, 37 (4), 397–409.

Randhawa, Ramandeep S. and Sunil Kumar (2008), "Usage restriction and subscription services: Operational benefits with rational users," *Manufacturing and Service Operations Management*, 10 (3), 429–447.

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70 (1), 41–55.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), pp.999–1033.

——— (1994), "Structual estimation of Markov decision processes," *Handbook of Econometrics*, 4 (51).

Schmalensee, Richard (1974), "Brand Loyalty and Barriers to Entry," *Southern Economic Journal*, 40 (4), 579.

Sinha, Jay I., Thomas Foscht, and Thomas T. Fung (2016), "How Analytics and AI Are Driving the Subscription E-Commerce Phenomenon," *Mit Sloan Management Review*.

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.

Tian, Longxiu and Fred M Feinberg (2020), "Optimizing price menus for duration discounts: A subscription selectivity field experiment," *Marketing Science*, 39 (6), 1181–1198.

Wager, Stefan and Susan Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113 (523), 1228–1242.

Xu, Kaiquan, Jason Chan, Anindya Ghose, and Sang Pil Han (2017), "Battle of the Channels: The Impact of Tablets on Digital Commerce," *Management Science*, 63 (5), 1469–1492.

Zhou, Jidong (2014), "Multiproduct search and the joint search effect," *American Economic Review*, 104 (9), 2918–2939.

Appendix A

Tables and Figures

Table A.1: Relationship Between Two Essays

	Effects on	
Stages	Focal Product	Other Products
Introduction	**Essay II**	
Subscribing		**Essay I**

Table A.2: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
weekly purchases					
spending	640,325	108.653	389.476	0.000	47,788.920
# of categories	640,325	1.006	1.782	0.000	13.000
# of products	640,325	2.926	6.268	0.000	112.000
# of trips	640,325	3.396	8.222	0.000	582.000
customer info					
gender (=1 if male)	3,659	0.581	0.493	0	1
referred (=1 if referred)	3,659	0.138	0.345	0	1
family size	3,659	2.991	0.741	1	9
sign-up date	3,659	8/30/2013		8/20/2009	10/1/2014

Table A.3: Summary of Subscription Products

	SKU Name	Starting in	Category	Frequency*
1	LEHO Reduced-fat Milk 1L	Feb, 2015	Dairy	2/week
2	FARMHOUSE Cage-free Eggs	Mar, 2015	ProcessedMeat&Eggs	1/week
3	ASAHI Milk 250ML	May, 2015	Dairy	1/day
4	HUNIING Eggs (6 Eggs)	May, 2015	ProcessedMeat&Eggs	1/week
5	COURTYARD Green Shell Eggs (8 Eggs)	May, 2015	ProcessedMeat&Eggs	1/week
6	FARMHOUSE Green Shell Eggs	May, 2015	ProcessedMeat&Eggs	1/week
7	WUCHANG Organic Rice	May, 2015	Grain&Miscellaneous	1/week
8	PLATEAU Rapeseed Oil	May, 2015	Grain&Miscellaneous	1/two-week
9	LEHO Reduced-fat Milk 200ML	May, 2015	Dairy	2/week
10	LEHO Whole-fat Milk 1L	May, 2015	Dairy	2/week
11	ASAHI Milk 950ml	Aug, 2015	Dairy	2/week
12	DONGLING Eggs (8 Eggs)	Dec, 2015	ProcessedMeat&Eggs	1/week
13	LEHO Reduced-fat Milk 200ML (3 pack)	Jan, 2016	Dairy	2/week
14	LEHO Whole-fat Milk 200ML (3 pack)	Jan, 2016	Dairy	1/two-day
15	COURTYARD Green Shell Eggs (12 eggs)	Apr, 2016	ProcessedMeat&Eggs	1/week
16	HUNING Eggs (12 eggs)	Apr, 2016	ProcessedMeat&Eggs	1/week
17	FARMHOUSE Rapeseed oil	Jun, 2016	Grain&Miscellaneous	1/two-week
18	EMEI Bottled Water 400ML (12 pack)	Aug, 2016	Beverages	1/week
19	EMEI Bottled Water 5L	Aug, 2016	Beverages	2/week
20	SIRMA Mineral Bottled Water (12 pack)	May, 2017	Beverages	1/week
21	WUMA Pure Milk	Jun, 2017	Dairy	1/week
22	JOKUL Milk 1L	Sep, 2017	Dairy	1/week
23	FARMHOUSE Rice	Sep, 2017	Grain&Miscellaneous	1/week

Table A.4: Differences of Customers in Treatment and Control in First Three Months

	Mean of Control	Mean of Treatment	p value	t-statistic
#of categories	1.847	3.035	0	-20.444
#of products	5.853	11.467	0	-14.058
spending($)	221.963	397.005	0	-12.920
#of categories in web	0.654	1.253	0	-8.771
#of products in web	1.977	4.267	0	-8.853
spending in web	60.820	125.954	0	-7.043
#of categories in tel	1.350	2.262	0	-15.431
#of products in tel	4.205	8.103	0	-10.164
spending in tel	185.493	335.499	0	-11.164
#of categories in app	0.281	0.621	0	-7.202
#of products in app	0.768	1.826	0	-6.952
spending in app	24.381	55.590	0	-6.158
#of categories in offline	0.122	0.299	0.0002	-3.753
#of products in offline	0.545	1.506	0.001	-3.372
spending in offline	21.312	62.490	0.001	-3.352
recency	-1.981	-1.315	0	-6.767
frequency	5.083	12.851	0	-18.868
$ in Beverages	0.272	0.164	0.565	0.575
$ in Dairy	55.359	196.455	0	-7.092
$ in Fresh meats	356.264	873.583	0	-11.607
$ in Fruits	217.611	555.680	0	-9.481
$ in Gifts	48.566	165.982	0.00000	-4.637
$ in Grain&miscellaneous	53.480	195.603	0	-10.494
$ in ProcessedMeat&Eggs	41.668	133.102	0	-10.925
$ in ReadyToEat	150.144	427.286	0	-12.726
$ in Seafood	39.784	94.678	0	-6.900
$ in Seasoning	21.075	61.388	0	-6.389
$ in Snacks	32.263	89.599	0	-7.645
$ in Vegetable	221.736	620.077	0	-11.819
gender	0.617	0.485	0	6.952
sign-up week	-69.562	-73.015	0.040	2.052
referred	0.135	0.137	0.926	-0.093
$ on FARMHOUSE Eggs	160.080	105.824	0	14.183
$ on WUCHANG Rice	196.968	191.465	0.0002	3.754
$ on PLATEAU Rapeseed Oil	199.000	196.129	0.003	2.975
$ on ASAHI Milk 250 ml	196.737	187.879	0.00000	5.166
$ on ASAHI Milk 950 ml	193.328	178.659	0	6.561
$ on LEHO Whole-fat Milk 1L	196.749	190.758	0.0001	3.936
$ on LEHO Reduced-fat Milk 1L	197.510	193.344	0.001	3.192

Table A.5: Post-match Differences of Customers in Treatment and Control

	Mean of Control	Mean of Treatment	p value	t-statistic
#of categories	2.764	2.764	0.998	0.003
#of products	9.632	9.607	0.952	0.060
spending($)	342.099	337.131	0.750	0.318
#of categories in web	1.031	1.113	0.375	-0.888
#of products in web	3.229	3.559	0.318	-0.999
spending in web	98.128	104.989	0.496	-0.681
#of categories in tel	2.061	2.035	0.751	0.317
#of products in tel	6.998	6.791	0.600	0.525
spending in tel	291.340	284.163	0.661	0.439
#of categories in app	0.572	0.553	0.772	0.290
#of products in app	1.594	1.531	0.748	0.322
spending in app	48.355	48.360	0.999	-0.001
#of categories in offline	0.228	0.224	0.943	0.072
#of products in offline	1.108	0.983	0.706	0.378
spending in offline	43.868	37.784	0.645	0.461
recency	-1.338	-1.543	0.105	1.621
freq	10.052	10.086	0.947	-0.066
$ in Beverages	0	0.199	0.206	-1.265
$ in Dairy	129.778	132.332	0.897	-0.130
$ in FreshMeats	693.870	697.557	0.944	-0.070
$ in Fruits	421.029	425.225	0.914	-0.108
$ in Gifts	97.801	92.081	0.781	0.278
$ in Grain&Miscellaneous	135.551	135.727	0.990	-0.012
$ in ProcessedMeat&Eggs	93.207	99.934	0.494	-0.684
$ in ReadyToEat	324.816	323.973	0.976	0.030
$ in Seafood	82.872	75.228	0.449	0.756
$ in Seasoning	45.229	41.211	0.449	0.757
$ in Snacks	66.159	69.441	0.747	-0.323
$ in Vegetable	474.676	467.425	0.868	0.166
gender	0.511	0.512	0.958	-0.053
sign-up week	-73.221	-72.778	0.852	-0.187
referred	0.120	0.138	0.309	-1.018
$ on FARMHOUSE Eggs	118.175	118.680	0.925	-0.094
$ on WUCHANG Rice	193.441	193.083	0.853	0.185
$ on PLATEAU Rapeseed Oil	197.514	198.617	0.284	-1.072
$ on ASAHI Milk 250 ml	191.660	191.634	0.990	0.012
$ on ASAHI Milk 950 ml	185.565	185.039	0.851	0.187
$ on LEHO Whole-fat Milk 1L	193.115	191.359	0.398	0.846
$ on LEHO Reduced-fat Milk 1L	194.809	194.517	0.866	0.169

Table A.6: The Effects of Subscription on Other Products

	Dependent variable:		
	DID + Static PSM	DID + Dynamic PSM	DID + LA PSM
	(1)	(2)	(3)
Subscription	0.670***	0.508***	0.244**
	(0.047)	(0.043)	(0.078)
Observations	227,654	236,486	48,720
R^2	0.406	0.384	0.392
Adjusted R^2	0.401	0.374	0.383
Residual Std. Error	2.118 (df = 226079)	2.207 (df = 232823)	2.212 (df = 48031)
Note:			. p <0.1; * p <0.05; ** p <0.01; *** p <0.001

Table A.7: Placebo Tests Results for Three Matching Methods

	Dependent variable:		
	Static PSM	Dynamic PSM	LA PSM
	(1)	(2)	(3)
treated_fake	0.102	0.075	0.072
	(0.062)	(0.056)	(0.102)
Observations	48,215	56,934	25,920
R^2	0.367	0.394	0.416
Adjusted R^2	0.355	0.383	0.400
Residual Std. Error	2.252 (df = 47331)	2.216 (df = 55896)	2.197 (df = 25229)
Note:			. p <0.1; * p <0.05; ** p <0.01; *** p <0.001

Table A.8: Evidence for the Reminding Effect of Subscriptions: Effects on the Probability of Purchase During Delivery Day

	Dependent variable:
	Delievery as Reminder
Subscription	0.029***
	(0.004)
Subscription*Delivery	0.601***
	(0.016)
Observations	3,944,612
R^2	0.266
Adjusted R^2	0.265
Residual Std. Error	0.224 (df = 3939983)
Note:	. p <0.1; * p <0.05; ** p <0.01; *** p <0.001

Table A.9: Five Most Frequent Rules

Association Rule	Support	Confidence
{item1 = chicken egg} => {item2 = sugar}	3.94%	22.21%
{item1 = chicken egg} => {item2 = ketchup}	3.00%	32.34%
{item1 = chicken egg} => {item2 = green onion}	2.55%	46.91%
{item1 = chicken egg} => {item2 = tomato}	2.17%	32.53%
{item1 = chicken egg} => {item2 = minced pork}	1.86%	83.57%

Table A.10: Evidence for the Reminding Effect of Subscriptions: Effects on the Purchases of Combo and Non-combo Products

	Dependent variable:	
	Complementary Products	Non-complementary Products
	(1)	(2)
treated	0.514***	0.430***
	(0.038)	(0.042)
Observations	351,946	428,099
R^2	0.365	0.495
Adjusted R^2	0.361	0.492
Residual Std. Error	1.750 (df = 349599)	1.770 (df = 425279)
Note:		. p <0.1; * p <0.05; ** p <0.01; *** p <0.001

Figure A.1: Illustration of the Subscription Plans of Eggs

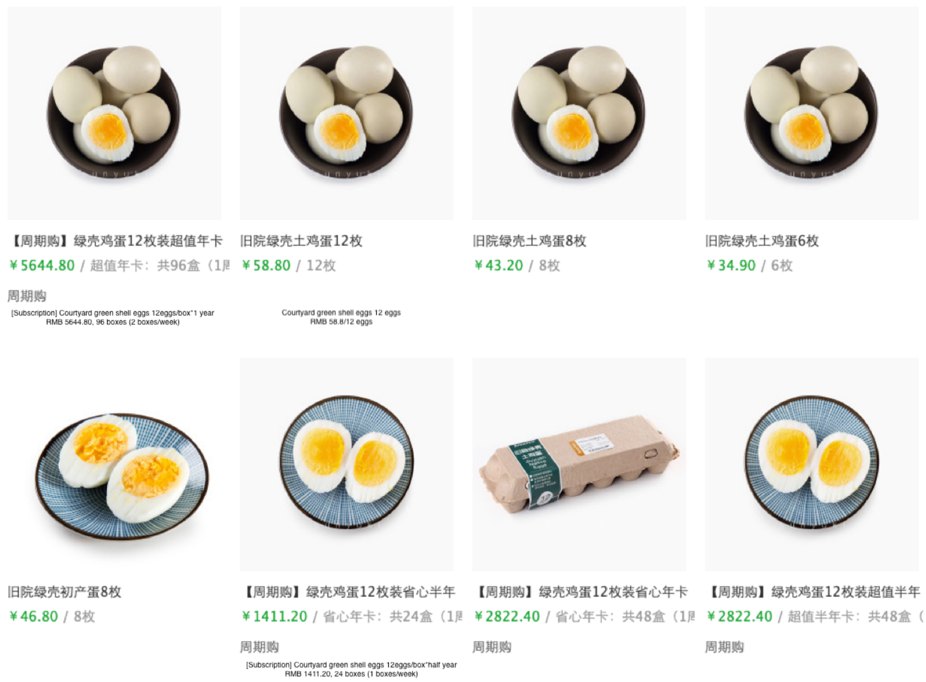

Figure A.2: Illustration of the Three Matching Methods

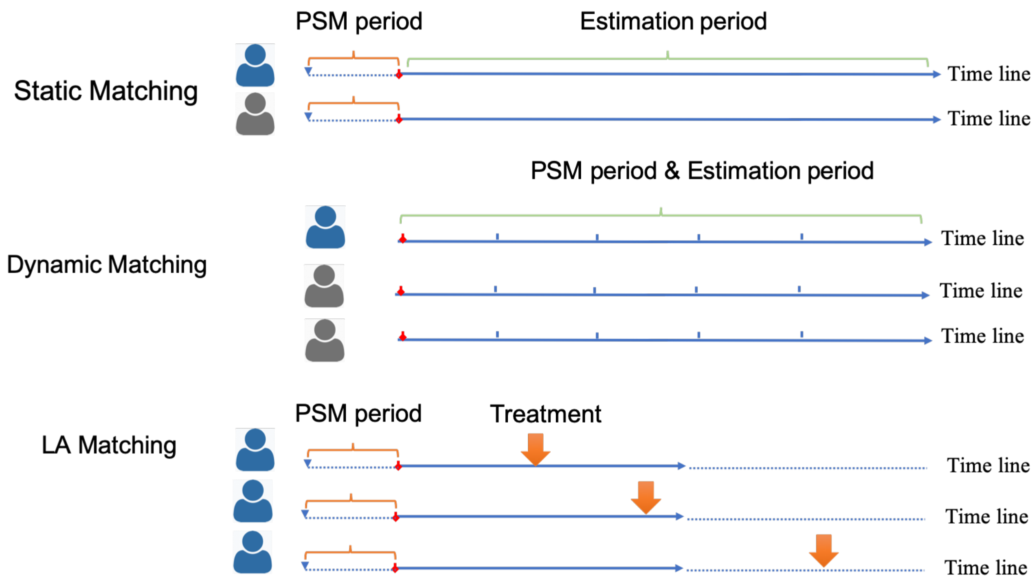

Figure A.3: Illustration of the Three Matching Methods

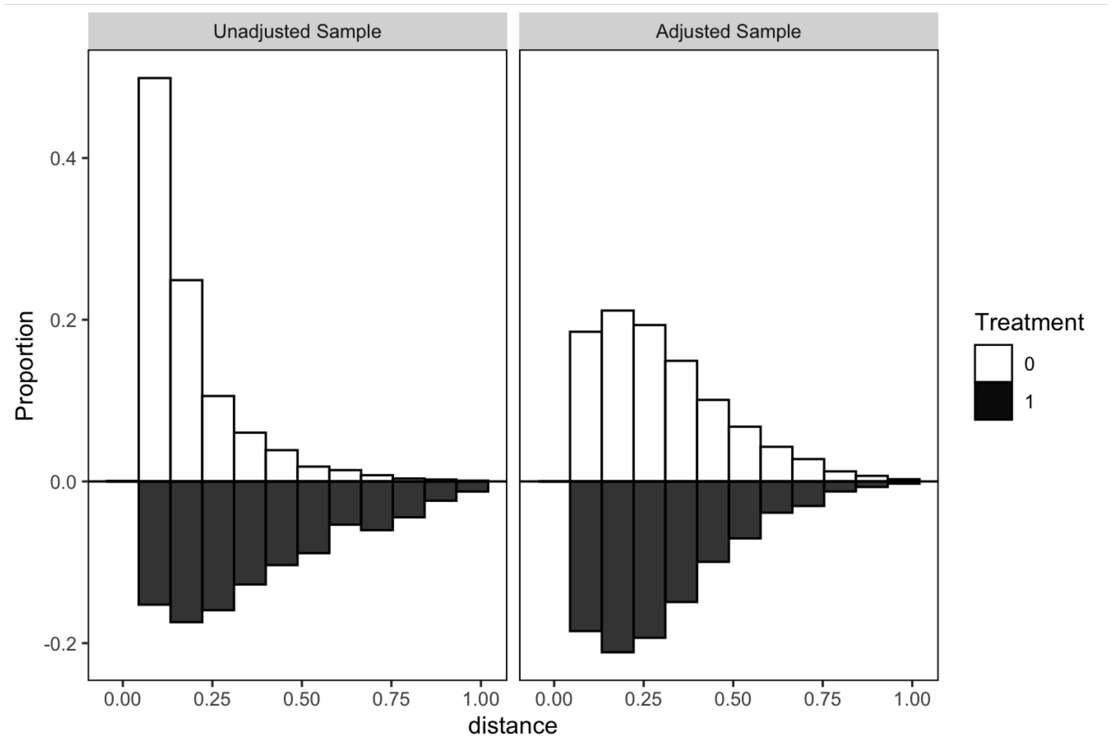

Figure A.4: Category-Level Effects of Subscriptions

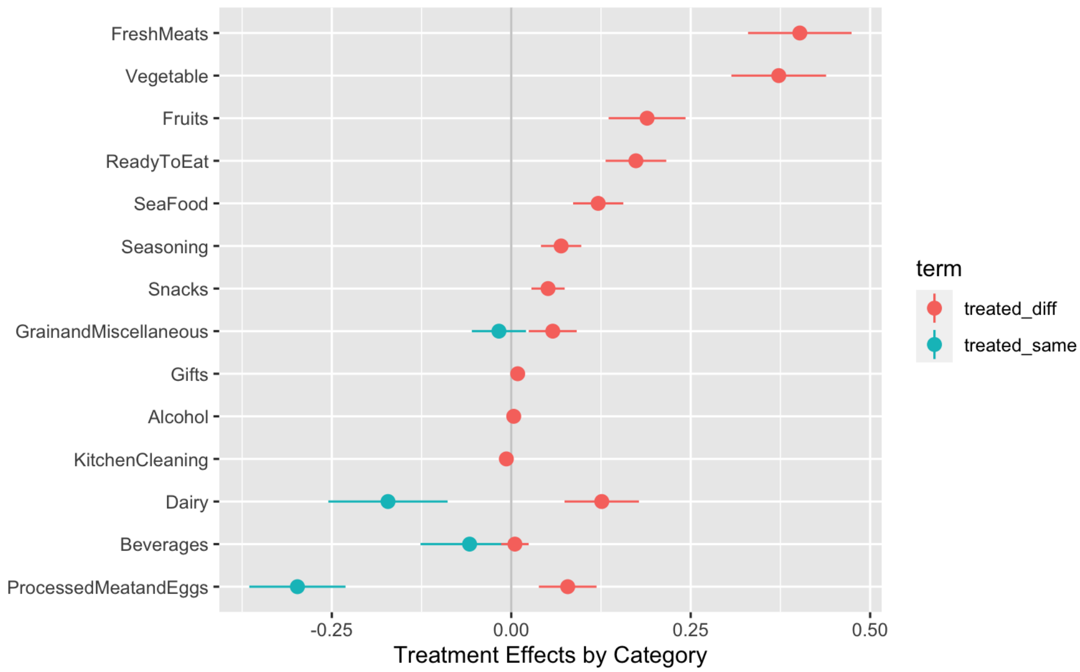

Figure A.5: Recipe Example

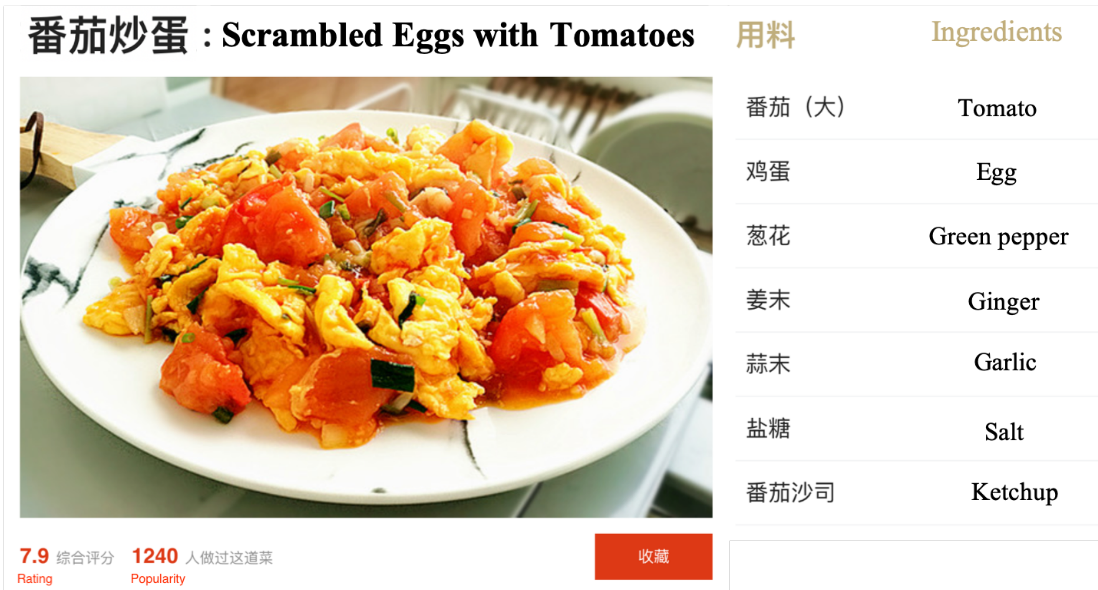

Figure A.6: Illustration for the Apriori Algorithm

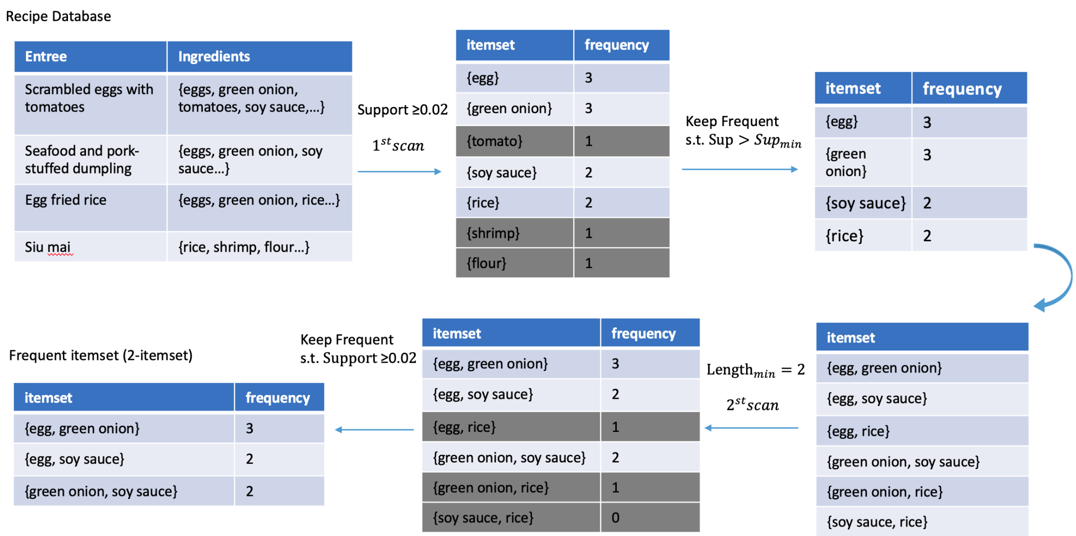

Table A.11: Step 1 Results - Brand Choice

	Dependent variable:		
	choice		
	(1)	(2)	(3)
total_price	-0.0127^{**}	-0.0342^{**}	-0.0241^{*}
	(0.0046)	(0.0113)	(0.0116)
total_price*gender		0.0508^{***}	0.0433^{**}
		(0.0154)	(0.0158)
total_price*new		-0.0032	-0.0062
		(0.0195)	(0.0211)
total_price*large		-0.0073	-0.0132
		(0.0183)	(0.0197)
brand fixed effect			✓

Note: . p <0.1; * p <0.05; ** p <0.01; *** p <0.001

Table A.12: Step 3 Results - Size Choice

	Family Size ≤ 3	Family Size > 3
Utility - linear	141.939	138.389
Utility - quadratic	-0.462	-0.545
Transaction Cost	3.771	3.790
Log likelihood	-4733.801	-1075.170

Table A.13: Size Choice Comparisons

	Family Size ≤ 3		Family Size > 3	
	Hybrid	One-time Only	Hybrid	One-time Only
0 eggs	4275	4270	1075	1072
6 eggs	297	184	64	44
12 eggs	441	238	117	55
18 eggs	235	293	5	66
24 eggs	60	430	8	93
12 eggs, 12 weeks	59		47	
24 eggs, 12 weeks	48		14	
Total Number of Eggs	35064	19557	12870	4346
Ratio	0.558		0.338	
Overall Ratio	0.499			

Figure A.7: Prices of Different SKUs of Chicken Eggs