# Essays on Dynamic Consumers' Brand Choice

Nahyeon Bak

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Pat Bajari, Chair

Chris Anderson

Dong-Jae Eun

Yuya Takahashi

Program Authorized to Offer Degree:
Department of Economics

University of Washington

**Abstract**

Essays on Dynamic Consumers' Brand Choice

Nahyeon Bak

Chair of the Supervisory Committee:
Title of Chair Pat Bajari
Department of Chair

This dissertation is a collection of essays on consumer's state dependent choice. In many consumers packaged goods markets, consumer's brand choice is highly persistent because of state dependence where past choice directly influence present choice. Chapter I investigates why consumer choices show state dependence by testing two competing theories: learning and switching costs. To test them, I used a Nielsen consumer panel data set including a long history of repeated purchases by 28,724 households from 2006-2015. Reduced form estimates suggest that the results align with learning, but not switching costs. I also find the only the first and second brand experiences affect present choice.

In Chapter II, consistent with reduced-form analysis, I hypothesize that under learning behavior, if consumers try a new brand, consumers are likely to choose a smaller size than before because of uncertainty on product information, if not, consumers are likely to choose a bigger size than before because of lower price per unit with a bigger size. However, under switching cost behavior, consumers size choice will not be affected by brand switching decision. To test this causal relationship between brand switching decision and size choice, I adopt double machine learning method. Compared to previous reduced-form analysis, double machine learning model specifies a set of control variables without human judgement and it provides a causal parameter. Also, compared to naive or prediction based machine learning models, it overcomes the regularization bias by using Neyman orthogonality and over-fitting

problems by using sample splitting method. As a result, I find that consumer's new trial on a brand leads to choose a smaller size choice than before where it supports learning behavior, not switching costs behavior.

These reduced form studies of Chapter I and II motivate structural approaches to empirical modeling. Chapter III tests the two competing theories with a structural demand model that incorporated variety-seeking behavior. Previous studies failed to explain how states affect two decisions: not only persistent brand choice, but also brand switching that usually variety-seeker have shown. To incorporate these decisions, I develop a dynamic panel demand model with multiple discreteness choices for estimating preferences where some consumers switch brand frequently even most consumers show persistent brand choice. I first find that consumers learn fast, which disputes previous slowdown learning models such as Bayesian learning. Second, state dependence of consumer choice diminishes with time elapsed from each purchase. These findings are robust to controlling variety seeking behavior or not.

Combining Chapter I, II, and III, I conclude that with the assumption on myopic consumers, because of learning behavior, consumers show persistent brand choice in the initial shopping period, but as they exposure to the same brands again and again, they become satiated the brand. In other words, consumers show diminishing marginal utility over quantity consumed. Therefore, consumers switch a brand.

# TABLE OF CONTENTS

# LIST OF FIGURES

# ACKNOWLEDGMENTS

First and foremost, I thank Pat Bajari for graciously serving as my advisor. He provided me with wholehearted comments to motivate and develop my research. His direction on training machine learning widened my scope of research and view. Also, his insight into the research guides me to improve the quality of the dissertation significantly. Furthermore, continuous support and encouragement presented me with a sustainable Ph.D. life. Additionally, I thank Patrick Bajari for his extensive and conscientious advising and professional assistance, which far exceeded his formal role as a dissertation committee member.

I also thank committee members, Chris Anderson and Dong-Jae Eun, and Yuya Takahashi. Chris Anderson was always generous with his time. He patiently offered constructive, insightful feedback on the development of my research. I also thank Dong-Jae Eun and Yuya Takahashi, for their valuable comments, advice, and time on my dissertation.

My research also benefited greatly from the support of faculty members Yanqin Fan, and Fahad Fahad Khalil.

# DEDICATION

To my dear husband and daughter, Daisoon and Seoyoon

# Chapter 1

# SOURCES OF DYNAMIC CONSUMERS' BRAND CHOICE

## *1.1 Introduction*

Consumers show persistent brand choices for consumer-packaged goods. Prior research explains that this persistence is due to structural state dependence where past choices directly affect present choice, even accounting for unobserved consumer differences. The fields of economics and marketing have reached a consensus on consumers' choice dynamics, especially on state dependence. However, a more important question, the main mechanism that generates state dependence, still needs to be answered. As the main source of state dependence, there are two competing theories: learning and switching costs. Identifying the true source of state dependence and quantifying it are important for firms' optimal policy and measuring social welfare. Each explanation implies different managerial policies and optimal pricing policies for firms, such as pricing, discounts, new product releases, expenditure on advertising, and research and development.

The primary goal of this article is to empirically identify the true source of state dependence by testing two competing theories: switching costs and learning, as suggested by [42], [81], [33], [76], and [71].[1] In this paper, according to [64], uncertainty about the quality of untested brands(Informational investment) causes persistent choice in the learning model, and the source of persistent choice in switching cost model is consumers psychological costs of switching, or non-economic "brand-loyalty" (Psychological investment). First, the switching costs model suggests that past purchases increase the current utility derived from the con-

[1] [54] suggest stockpiling behavior as one of the sources generating dynamics. It explains that consumers buy more during promotions not only for current use but also for future consumption by stockpiling. Also, the key assumption of stockpiling behavior is a forward-looking consumer. I do not find any evidence of stockpiling in this data.

sumption of the product, and the consumer will not get this increased utility if the consumer switches brands. Thus, this increased utility plays the role of fixed switching costs.[2] On the other hand, in learning behavior, consumers are uncertain about unobserved characteristics such as quality or true flavor, and by experiencing the product, they update their information and reduce their uncertainty. Thus, the cost of uncertainty subsequently decreases following experience.[3]

Each theory provides different implications for both decisions: persistent brand choice and brand switching. In the cough drop market, firms supply various sizes of products, and the bigger the consumer's size choice is, the lower price per count consumers pay. According to the learning model, when a consumer chooses an experienced brand, they are more likely to choose a bigger size than before due to the reduced cost of uncertainty, and a smaller size when they switch a brand. In contrast, the switching costs model implies that consumers always face constant switching costs, and there will be no substantial changes in size choice. Also, if consumers switch a brand due to a temporary price cut, the switching costs model expects that they will continue to choose the brand switched, but a learning model expects that consumers will return to the brand chosen previously. I implement tests that exploit these key differences in the predictions of the learning model and the switching costs model

[2]Here, the switching costs model includes habit-persistence, and the brand inertia(inattention) model in the perspective of econometric specification and behavioral implications.

[3]Following the broad definition of learning, here, I test whether experiences change beliefs over time. In other words, I test whether the probability of choosing a product changes as consumers become experienced. In the narrower definition of learning, consumers' choices will converge to some product as their posterior beliefs on product quality converge to some level under a stationary market environment. However, the market environment always changes, and consumer-packaged goods are horizontally differentiated, not only vertically differentiated. Therefore, it is difficult to observe the convergence in the data, even though consumers' learning exists. Also, if consumers are forward-looking, they will show strategic trials or experiments in the initial shopping period for future value [23, 27, 65]. However, here, I assume a myopic consumer. That is because I have not observed that consumers experiment with various products for future consumption, and the consumer does not expect when they will need a cough drop. For example, observations show that 19.69% of the households had purchased multiple items at a time. 8% of households switch brands at least once and experience at least two brands in initial shopping trips, and 82.26% of the households did not try other brands in the first three shopping trips. Also, the market share of new goods is under 0.05 percent. Consumers do not show strategic trials. [65] also shows that a myopic learning strategy may suffice as the magnitude of the utility shocks increases relative to the consumer's posterior uncertainty in quality.

and find little evidence in favor of the switching costs model. I conclude that the source of dynamics in my data is learning behavior, not switching costs behavior.

There have been some attempts to investigate sources of dynamics in brand choice, but researchers have not fully considered all possible sources of dynamics, such as satiation in variety-seeking behavior: a diminishing marginal utility as the level of consumption increases.[4] Moreover, previous studies have failed to explain the time-varying effects of state dependence. Also, in this research, providing evidence on structural state dependence is essential because unobserved consumer differences or auto-correlated taste shocks could lead to persistent brand choice, as well.

In this paper, I propose a new approach to overcome these challenges. I do this with three steps. The starting point for my analysis is to build a structural dynamic panel model incorporating variety-seekers who purchase multiple items at a time or frequently switch products. Based on a multiple-discrete choice model where a consumer $i$ 's utility at shopping trip $t$ is a sum of utility from each product purchased, suggested by [61], I control for variety-seeking behavior which also generates dynamic demand. Secondly, I specify a dynamic panel model that tests the two competing theories by allowing demand parameters to vary by past choices and the time elapsed from each purchase. The variables of time elapsed from each purchase measure the decaying effect of state dependence, which is usually ignored in previous research. Thirdly, I estimate utility parameters at the consumer level by applying a Bayesian Markov Chain Monte Carlo method which uses the Metropolis-Hasting(M-H) algorithm. This method makes inference about the posterior distribution of heterogeneous model parameters across consumers or products.

To estimate my dynamic panel demand model, I use a rich data set from a Nielsen consumer panel that tracks the majority of shopping trips for packaged goods by 28,724

[4] I identified that 39% of households are variety-seekers who purchase multiple items at a time or frequently switch products. This variety-seeking behavior could be a source of dynamics, but it is unable to explain all consumers' brand choices. Furthermore, variety-seeking behavior and learning behavior could coexist. Therefore, I only test two competing theories: the learning model and switching costs model by controlling for variety-seeking behavior with a multiple-discrete choice model

households in the US between 2006 and 2015.[5] This unique data set helps to estimate high-dimensional dynamic demand and to investigate sources that generate dynamics, such as switching costs and learning while controlling for the effects of unobserved consumers' differences and variety-seeking behavior. I apply my model to the data on a variety of cough drop purchases. Later, as a robustness check, I test two theories for other consumer-packaged goods, including yogurt and margarine. In the data set for the cough drop, 11% of the transactions of cough drops involve a purchase of multiple items, and the probability of brand switching is 0.5 at the 90th percentile of consumers. This simultaneous and sequential demand for varieties requires me to use a multiple-discrete choice model with a satiation parameter.

I find evidence of structural state dependence, which is robust to controls for unobserved consumer differences. The estimation results indicate that the main source of generated dynamics is learning, but only the first two purchases of a brand affect current brand choice. In contrast to a slowdown learning model such as Bayesian learning, the dependence on the first and second choice of a brand shows that consumers learn fast. In addition, the model incorporating satiation behavior also explains data better than the model ignoring it.

My model and estimator build upon the structural state dependence literature in several ways. First, I develop a novel approach for identifying state dependence, which has long been a thorny issue in the literature on serially correlated dynamic demand. My approach is a more general thought of state dependence model incorporating competing models. Also, I allow for a rich form of heterogeneity by using a Bayes hierarchical model and addressing the role of unobserved heterogeneity in state dependence. The suggested method and identification strategy are quite general and can be applied whenever it is important to distinguish state dependence and unobserved heterogeneity. Second, to the best of my knowledge, I first

[5] Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

examine these two representative competing models jointly, and my results suggest that the best explanation for the source of dynamics in consumer demand is fast learning, whereby only the first and second brand experiences change the conditional repurchase probability for the brand during the future shopping trip. I refer to fast learning as Non-Bayesian learning, hereafter. My specification includes the decaying effect of state dependence, in contrast to previous literature which concludes that the source of dynamics is switching costs. Hence, I estimate unbiased state dependence terms, and they explain the time-varying effect of state dependence.[6] Third, based on not only the estimation results but implications from each behavioral mechanism, I conclude that the consumer demonstrates non-Bayesian learning behavior.

The remainder of the paper proceeds as follows: in Section 2, I introduce a structural dynamic demand panel model, which identifies structural state dependence and tests two competing theories to find the source of structural state-dependence; Section 3 explains estimation; Section 4 describes the main components of the data set and identification; Section 5 demonstrates learning behavior with descriptive analysis and reduced-form evidence: estimates of the structural state dependence model and results are in Section 6; and Section 7 provides a conclusion.

### *1.1.1 Related Literature*

This study is related to the literature on (a) disentangle sources of dynamic choice between switching cost and learning, (b) fast learning from experience, and (c) satiation behavior.

First, many empirical papers have distinguished learning from dynamic mechanisms such as the switching costs model as a true source of state-dependence, including [72], [33], and [71]. However, few empirical papers estimate a structural model with both an initial condition and a decaying effect. If the model omitting these two effects, and consumers already have enough experiences of a brand, measuring the learning effect will not be possible. See [26, 69]

---

[6] Only [82] considers the decaying effect, but on a lagged choice, not all past choices. It also shows that the effect of past choice diminishes with time elapsed from the previous purchase.

for the details. Also, the specification of switching costs in the previous paper includes only lagged brand choice in the model( [82], [32], [31], [33]). It also represents the recency effect in learning, not only switching costs. By contrast, this paper aims to quantify the learning effect in a more general context of a structural dynamic demand model and shows that the main source of state-dependence is learning behavior, not switching costs with reduced-form evidence of size choice.

Second, this paper considers experiential learning, not observational learning, learning through search [75, 72], learning from advertising [1]. According to [40], use experience is the most accurate signal of quality than price signals and advertising signals. Most experiential learning models adopt Bayesian learning model, defined by a multi-armed bandit problem [30, 74, 76, 21, 26, 27, 22], or index strategies, defined by restless bandit problem [65]. With myopic consumer assumption[7], I propose a more generalized dynamic panel model while controlling for different learning rates across consumers. With this model, I show that consumers learn fast with 1-2 times purchase in contrast to slow down learning models such as Bayesian learning.

Third, to the best of my knowledge, this paper is the first literature integrating satiation behavior with learning behavior. Despite previous literature explaining satiation behavior, described by diminishing marginal utility, there were no studies considering that brand switching decision also depends on previous brand choice. See [67, 68, 7, 61, 50, 62, 52, 91]. By contrast, to fully understand the state-dependent brand choice, this paper explicitly controls for satiation behavior and assesses the learning behavior.

---

[7]In this data-set, I do not find evidence of forward-looking behavior. However, to understand forward-looking consumers' choice, Chapter 4 studies how forward-looking consumers' choices are different from myopic consumers

## 1.2 Identification and Data

I use a Nielsen consumer panel data on cough drop purchases with 28,724 households in the US between January 2, 2006, and December 26, 2015.[8] The cough drop category in this data set has a number of unique features to address the sources of dynamics. First, the panel data set allows one to keep track of brand choices made by each household over time and across stores. This helps to control for the effect of time elapsed from the previous purchase on brand choice and the effect of changing consideration set on brand choice.

Second, consumer behavior in the cough drop market helps to measure the effect of learning precisely. Since consumers cannot predict when they will need a cough drop, they do not show stockpiling behavior. Therefore, the timing of the purchase is likely to be consistent with the timing of consumption. Though only the date of purchase is observable, I am able to estimate the effect of information acquired from past consumption on brand choice. From this data, I observe that consumers tend to purchase cough drops when they are needed rather than to stockpile them for future use. This consumption behavior helps to exclude stock-piling behavior as one of the dynamic sources.

Third, consumers purchase cough drops infrequently but periodically because of the regular flu season. This infrequent purchase pattern helps to solve an initial condition problem that the transaction was not recorded in this data. By defining a new consumer as a consumer who made a purchase at least one year later than their last purchase, I control for the effect of past shopping history, which is not recorded.[9] Focusing on these new customers greatly alleviates the problem of initial conditions that all dynamic models are subject to, which is even more problematic for learning models. This strategy also guarantees high

[8]These data are available for academic research through a partnership with the Kilts Center at the University of Chicago, Booth School of Business. See http://research.chicagobooth.edu/nielsen for more details on the data.

[9]I observe that 71% of the households in the panel have purchased cough drops at least a year later from the previous purchase. Hence, I treat consumers who purchased cough drops at least a year later from the previous purchase as new consumers. That is because most consumers do not remember what they purchased a year ago and cough drops are less likely to be stored over a year because of their small size.

internal validity, as seen in Appendix Table A1.

Two national brands dominate the market during the data period: Halls and Ricola.[10] On average, the two brands' market share is 65% from 2007-2015. I define the outside goods in the cough drop category as any brands other than Halls and Ricola. Using this definition of the outside goods, I only consider those shopping trips where purchases in the cough drop category are made. After defining new consumers, I solely use those who have made at least four purchases over the study period. This yields a sample of 127,581 transactions by 28,724 households.[11]

### *1.3 Descriptive Analysis and Reduced Form Evidence*

This section provides evidence of learning in the data. Conceptually, switching costs and learning affect consumers' brand choices in different ways. In the switching costs model, past consumption increases the current utility derived from the consumption of the product. This increased utility plays a role of fixed cost in brand switching and leads to persistent brand choice. Therefore, the marginal purchase probability will be lower than the conditional repurchase probability for each of the product considered. Moreover, the conditional repurchase probability for the product should not change over experience.

A learning model also implies that past consumption increases the current utility derived from the consumption of the product. However, it is true on the condition that the utility from product characteristics experienced is not less than the expected utility from product characteristics.[12] As consumers get more experienced, the cost of uncertainty on product characteristics decreases, and the repurchase probability for the product increases. Therefore, if consumers show learning behavior, the lowest repurchase probability is after the first

[10] Hereafter, I treat the following four brands as one brand, Halls: Halls, Halls max, Halls warm-ups, and Halls naturals

[11] I run a structural model with the data from 10% of the random households who have purchased one of the top nine products from the two brands in the sample because it takes at least to run each model. It includes 2,098 transactions from 592 households.

[12] If the utility from characteristics experienced is less than the expected utility from characteristics at the first experience of a brand, a consumer is likely to switch to another brand in the next shopping trip.

purchase of the product. Also, as the cost of uncertainty decreases, consumers are likely to choose a bigger size because the price per count of a bigger size is lower than that of a small size.

The challenge of distinguishing between switching costs behavior and learning behavior is that neither is observed. I only observe past and present brand and size choices. Thus, to disentangle switching costs behavior and learning behavior, I first conduct a descriptive analysis. To highlight the potential role of learning behavior and switching costs behavior in brand choice, I present the results of a simple exercise where I look for preliminary, reduced-form evidence of learning behavior. First, I estimate the size transition process describing the relationship between a decision of brand switching and a state of size from a multinomial logit model, and I show that my result supports the learning model, not the switching costs model. Also, I provide evidence on variety-seeking seeking behavior and emphasize its need for control.

### *1.3.1 Descriptive Evidence*

First, I provide suggestive evidence on structural state dependence and the effect of time elapsed from previous purchases. Table 1.1 shows that the marginal purchase probability at the first purchase and that of the fourth purchase is considerably smaller than the conditional repurchase probability. Around 62 to 76 percent of brand choices across brands is persistent. It suggests the structural relationship between past choice and current choice. For the transaction by consumers who repurchase cough drops within 30 days, around 76 to 81 percent of brand choices across brands are persistent. This implies that the time elapsed from previous purchase weakens the persistent brand choice.

Second, I find suggestive evidence on learning behavior from the fraction of brand switching over experience. In Table 1.2 and Figure 1.1, the fraction of brand switching after the first purchase of Halls is 0.36, but it decreases to 0.22 after the following purchase of Halls. This decreasing fraction of brand switching patterns for cough drop purchasing history is also true for Ricola. The fraction of brand switching after the first purchase of Ricola is 0.61

compared to 0.25 in the following purchase. The highest fraction of brand switching at the first experience of a brand supports the learning model and not the switching costs model implying constant repurchase probability. Most consumers are likely to realize that true characteristics such as flavor or quality are significantly different from the expected characteristics when they experience the product at the first purchase. If consumers' value of experienced characteristics is lower than the value of expected characteristics, *ceteris paribus*, they are likely to switch to other products at the next shopping trip. Also, the most frequently purchased size during the first purchase is 30 drops, which is enough to determine the true quality or actual flavors. Consistent with the descriptive evidence, the results in Appendix Table A3 and the fourth column of A6 also show that the highest probability of brand switching is right after the first purchase of each brand.

Third, consumers' brand switching initiated by temporary price cuts also provides evidence that switching costs are not a true source of state dependence. When consumer's spell (brand) is initiated by a temporary price cut, according to the learning model, consumers will return to the previous brand choice or try another brand as an experiment in the next shopping trip. In contrast, according to the switching costs model, consumers will not switch the brand again in the next shopping trip. Conditional on a purchase with a temporary price cut, I observe that 2720 spells are initiated with this temporary price cut. Of these, 46 percent of consumers repeat purchases and 54 percent of consumers switch to another brand (previous brand 66 percent, other brands 34 percent) in the next purchase.[13] Compared to the repurchase rate of Halls, 0.77, in Table 1.1, the choice of brand switched by temporary price cut is not persistent. The same is true for Ricola brand products with the conditional repurchase probability of 0.24 when given a temporary price cut, compared to the repurchase rate of Ricola, 0.61. These patterns are not consistent with the switching costs model.

Fourth, I find evidence for satiation in variety-seeking behavior. Satiation behavior means

---

[13] I treat at least a 10 percent discount as a temporary price cut.

Table 1.1: Switching Matrix

| | Share at the first purchase | Switching rate to Halls | to Ricola | to Non-national brands | Share at the 4th purchase |
|---|---|---|---|---|---|
| Halls | 56% | 78% | 4% | 18% | 56% |
| Ricola | 7% | 25% | 62% | 13% | 9% |
| Non-national brands | 36% | 27% | 3% | 69% | 35% |

Note: These shares are calculated from households who purchased cough drops at least four times. The diagonal of switching rate represents the conditional repurchase probability for each of the brands considered.

that preference for a unit of a brand is at the highest when the quantity consumed is very small, and as the quantity consumed grows, preference per unit drops. Thus, satiation could lead to brand switching, and it could mitigate persistent brand choice. To find evidence of satiation behavior, I define a spell as a sequence of consecutive choices of the same brand. Table 1.2 shows that for the median household who purchased cough drops at least four times, they switched brands three times, and once they switch a brand they tend to purchase it two times consecutively. For example, the median consumer shows the following history of brand choices: A, A, A, B, B, A, and B. This switching behavior is consistent with the satiation in variety-seeking behavior. This switching pattern hints that satiation behavior could be a potential driving force of structural state dependence and requires me to control for variety-seeking behavior to identify the true source of state dependence.

### *1.3.2 Size Transition Process*

Evidence of learning behavior can be found when consumers' cost of uncertainty on a product decreases over experience. As the cost of uncertainty falls, a consumer's willingness to pay for the product will increase. Consumers will choose to purchase a larger size of the product, compared to the previous choice of size. On the other hand, consumers are likely to choose a smaller size when they try a new brand, owing to a cost of uncertainty. To test this implication, I define brand-level consumption experience as the cumulative quantity of

Table 1.2: Descriptive Statistics: Household Purchase Behavior

| Variables | Mean | Std. Dev. | 10p | 50p | 90p |
|---|---|---|---|---|---|
| For the households who went shopping at least 4 times (Obs. 97,254) | | | | | |
| Total number of purchases | 9.24 | 9.07 | 4 | 6 | 17 |
| Time elapsed from the previous purchase (Unit:date) | 252.95 | 331.79 | 9 | 110 | 685 |
| No. of spells per households | 3.22 | 2.66 | 1 | 3 | 6 |
| Spell length for all spells | 2.87 | 4.03 | 1 | 2 | 6 |
| Fraction of brand switching(FBS) | | | | | |
| Fraction of brand switching(FBS) per households | 0.31 | 0.27 | 0 | 0.3 | 0.5 |
| FBS after the first purchase of Halls | 0.36 | 0.48 | | | |
| FBS after non first purchase of Halls | 0.22 | 0.42 | | | |
| FBS after the first purchase of Ricola | 0.61 | 0.48 | | | |
| FBS after non first purchase of Ricola | 0.25 | 0.43 | | | |

Note: 8,141 households who went shopping at least four times. obs. 97,254

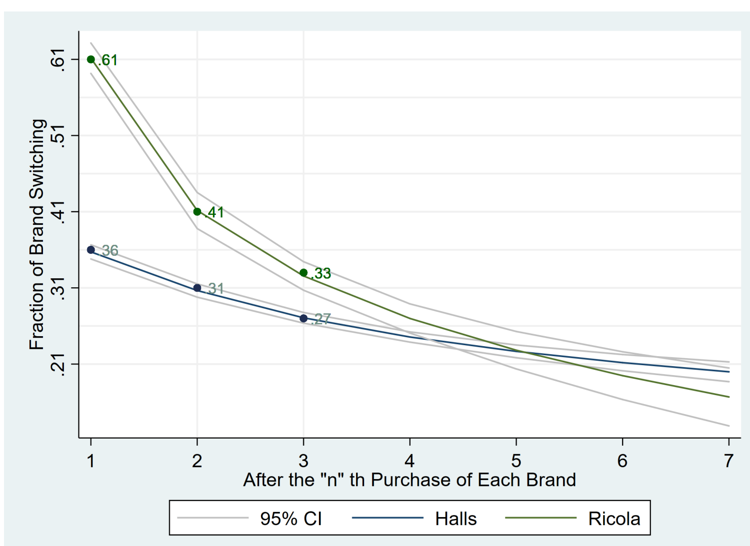



Figure 1.1: Fraction of Brand Switching over Experiences for Each Brand

purchases of the brand, $q_{i,t}$, and a decision on a new trial of a brand, $y_{i,t+1}$. I estimate the size transition process with a multinomial logit model from data, using consumption experience and a decision on a new trial:

$$\triangle size_{i,t} = \alpha_0 q_{i,t-1} + \alpha_1 y_{i,t} + \alpha_2 size_{i,t-1} + \alpha_3 y_{i,t} size_{i,t-1} + s'_{i,t}\beta + s'_{i,t-1}\gamma + \epsilon_{it} \tag{1.1}$$

where $s_i$ is a vector of states chosen by using the double selection method.[14] Also $\triangle size_{i,t}$, is a change of size at shopping trip t, and $size_{i,t-1}$ is a size choice at shopping trip t-1, respectively.[15]

Table 1.3 presents average marginal effects for a multinomial logit model representing the transition process. The average probability of choosing a larger size increases by 3.5 percent with a 1 percent increase in brand consumption, but the average probability of choosing a smaller size decreases by 5.5 percent. Also, the average marginal effect of the time elapsed from a previous purchase on choosing a larger size is -0.005, but on choosing a smaller size is 0.035. It shows that the probability of choosing a smaller size increases as the information decays, indicating that the cost of uncertainty increases as the time elapsed from previous purchase increases.

These results support the learning behavior related to decreasing cost of uncertainty. Also, the price elasticity of a brand decreases over experiences, reported in Appendix A Table A7. This diminishing price elasticity over experience reflects increasing willingness to pay caused by decreasing cost of uncertainty.

---

[14]First, I estimate choice probabilities using the regression of brand switching related to the decision on a new trial using machine learning (ML) methods such as random forest, Gradient Boost Machine, deep learning, and econometric models such as linear probability, logit, and probit. In the presence of high-dimensional parameters, machine learning methods provide important 183 variables among 1033 variables while controlling for a multi-collinearity problem. Because of limited interpretability of machine learning methods, I use econometric models with these 183 variables. The results are in Appendix Table A2 and A3. Thus, with these selected control variables, I run the main regression of size choice.

[15]These variables are defined in Appendix A.

Table 1.3: Average Marginal Effects for a Multinomial Logit Model

| | (1) | | (2) | |
|---|---|---|---|---|
| Variables | Larger size | Smaller size | Larger size | Smaller size |
| Log total counts accumulated | 0.035*** | -0.055*** | | |
| for the brand chosen$_{t-1}$ | (0.002) | (0.002) | | |
| Initial trials of a brand$_t$ | | | -0.035*** | 0.242*** |
| | | | (0.006) | (0.007) |
| I($size_{t-1} < 25$) | 0.397 | -2.462 | 0.562 | -2.738 |
| | (10.418) | (81.676) | (26.160) | (157.336) |
| I($49 < size_{t-1} < 99$) | -0.332*** | 0.288*** | -0.301*** | 0.282*** |
| | (0.004) | (0.003) | (0.003) | (0.003) |
| I($98 < size_{t-1}$) | -2.877 | 0.768*** | -3.175 | 0.859 |
| | (102.437) | (12.803) | (194.930) | (31.374) |
| Coupon Value$_t$ | 0.108*** | -0.105*** | 0.096*** | -0.10*** |
| | (0.004) | (0.006) | (0.004) | (0.005) |
| Log(date$_t$-date$_{t-1}$) | -0.005*** | 0.008*** | -0.010*** | 0.015*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |

Note: N=54,379. Significance levels: * p< 0.1, ** p<0.05, *** p< 0.01. The dependent variable is size choice and differentiates among three outcomes: a consumer choose (i) a smaller size than the previous choice, (ii) the same level of size, and (iii) a larger size than the previous choice.

The following control variables are included: interaction terms of size state with brand switching, deal, change of store, female head age, quarterly purchase frequency, seasonality, and trend.

## *1.4 Robustness Check: Learning Behavior in the Margarine and Yogurt Markets*

A final important question is whether my results can be generalized to other product markets. To show learning behavior as a source of dynamics, I test two product markets: margarine and Greek yogurt. In previous literature, [31, 33] shows that the source of consumers' inertia in their brand choice is a psychological switching cost in the margarine market, and [61] shows variety-seeking behavior in the yogurt market. For these two product markets, I estimate how the probability of brand switching changes over experiences and size transition process related to uncertainty cost, as seen in equation (11). Figures 1.2-1.5 show the results of estimations in the Margarine and Yogurt market. Consistent with the cough drop market, these two markets show that the highest fraction of brand switching is at the first experience of a brand. Also, they show that additional consumption increases the probability of choosing a larger size, but it decreases the probability of choosing a smaller size. In summary, these two results support learning behavior, rather than switching cost behavior.

## *1.5 Conclusion*

In this paper, I test whether the true source of state dependence in consumers' choice is learning behavior or switching costs in consumer-packaged goods market such as cough drops, margarine, and yogurt.

First, to distinguish between switching costs behaviors and learning behaviors, and to find the true source of state dependence, I characterize the difference between two behaviors in terms of definition and implication. In the switching costs model, since previous brand choice plays a role in fixed costs when switching to a brand, consumers show persistent brand choice. Also, since switching costs only depends on the previous brand choice, the conditional probability of repurchase for the product does not change over experience.

On the other hand, the learning model explains persistent brand choice from increasing utility over experience where past consumption increases the current utility derived from

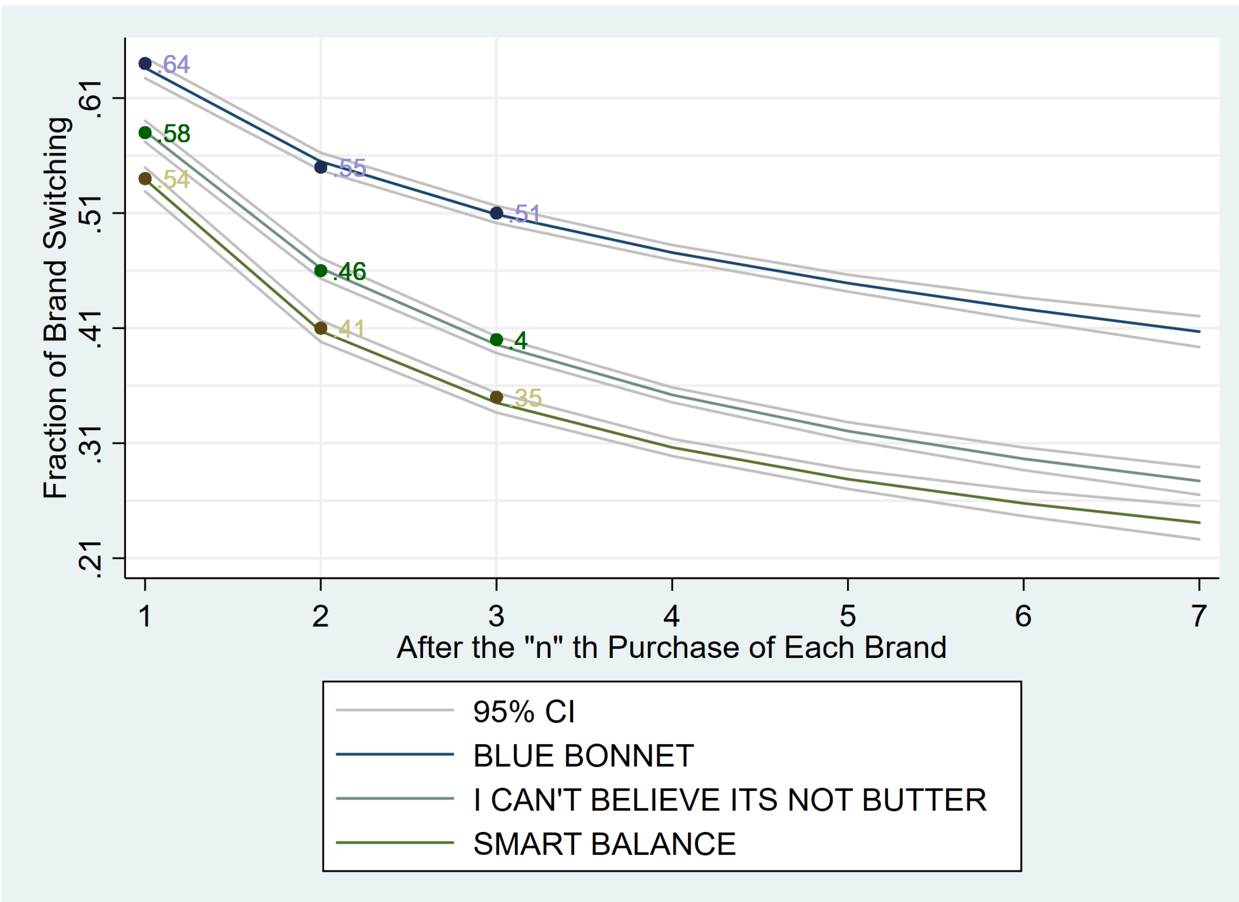



Figure 1.2: Reduced-Form Evidence in Margarine Market: Fraction of Brand Switching over Experiences for Each Brand

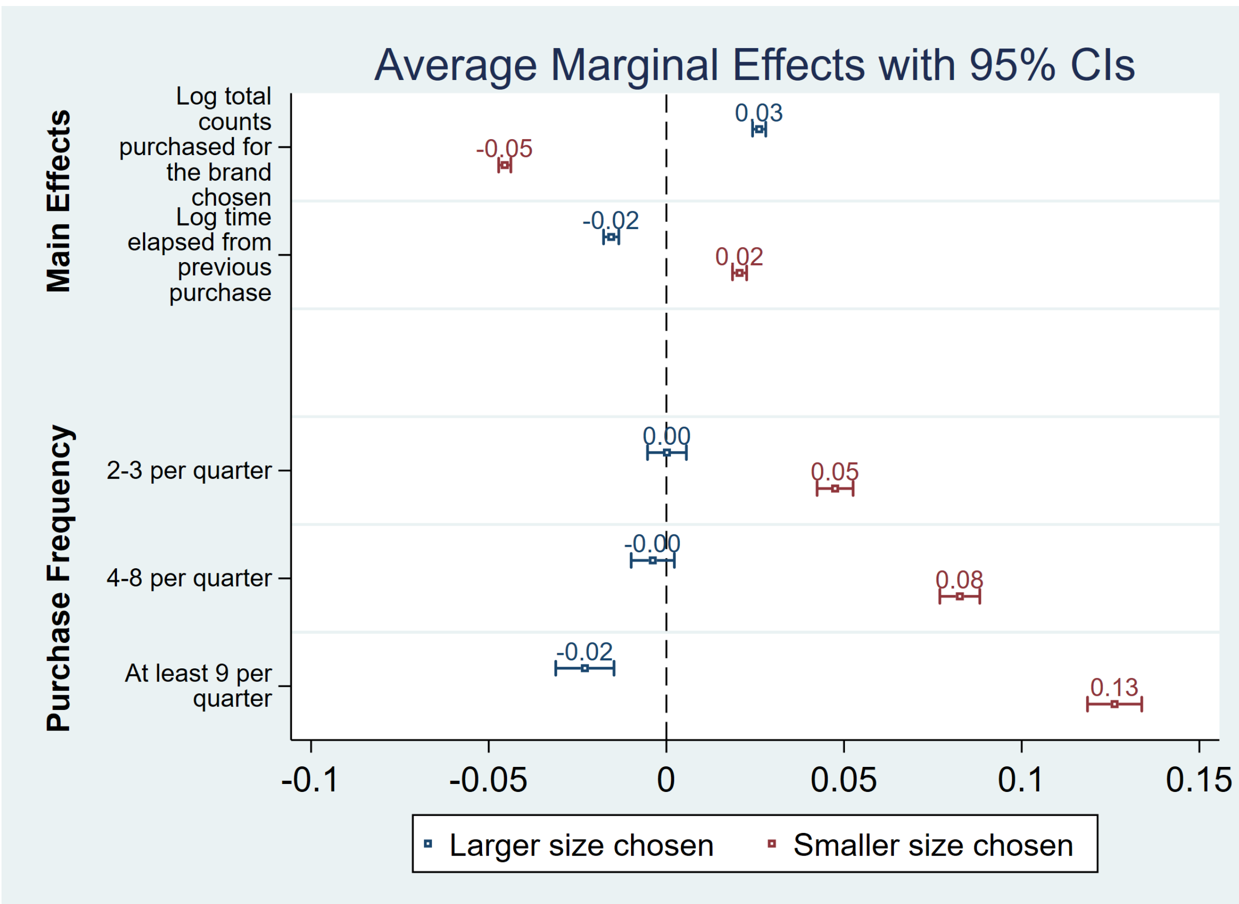



Figure 1.3: Reduced-Form Evidence in Margarine Market: Average Marginal Effects of Initial Trial of a Brand on Size Choice

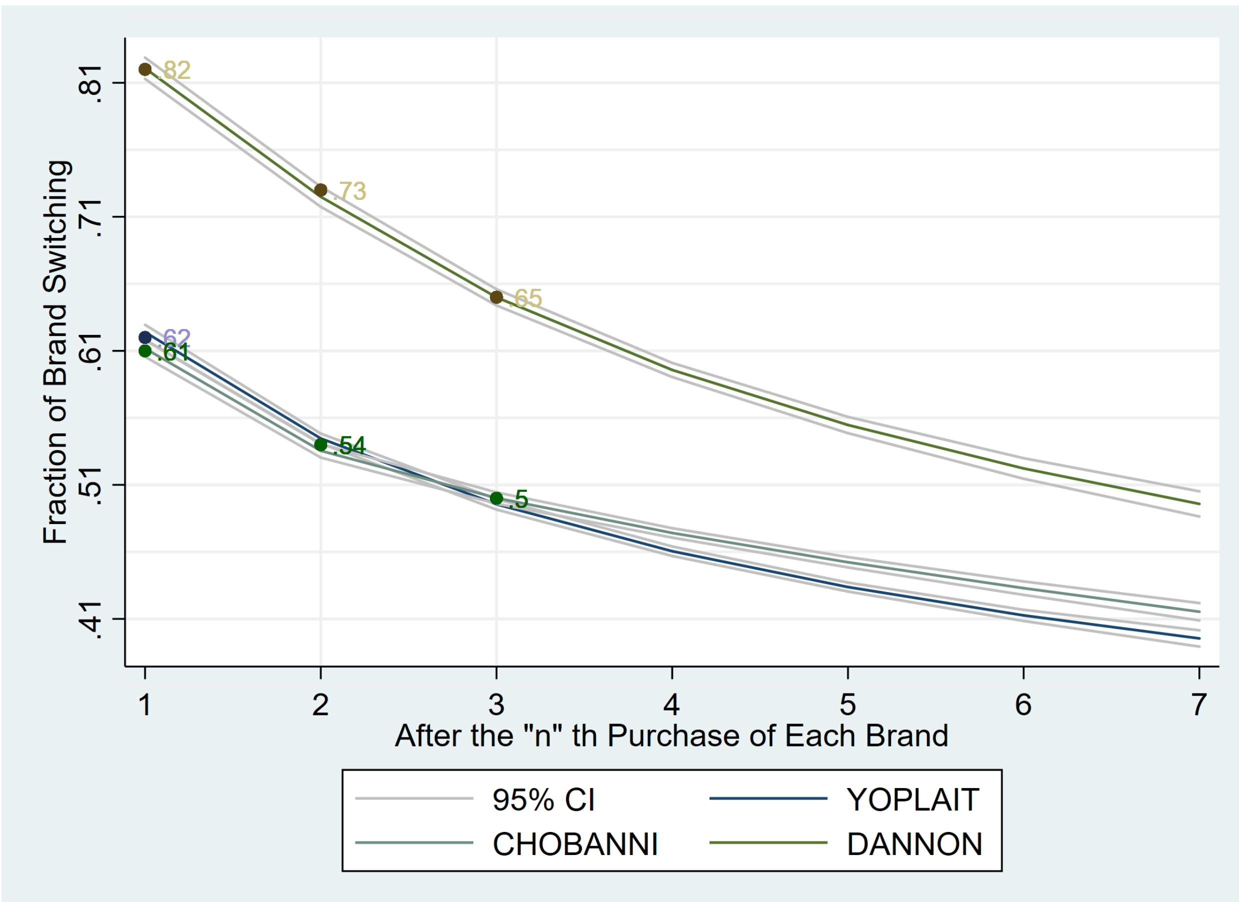



Figure 1.4: Reduced-Form Evidence in Yogurt Market: Fraction of Brand Switching over Experiences for Each Brand

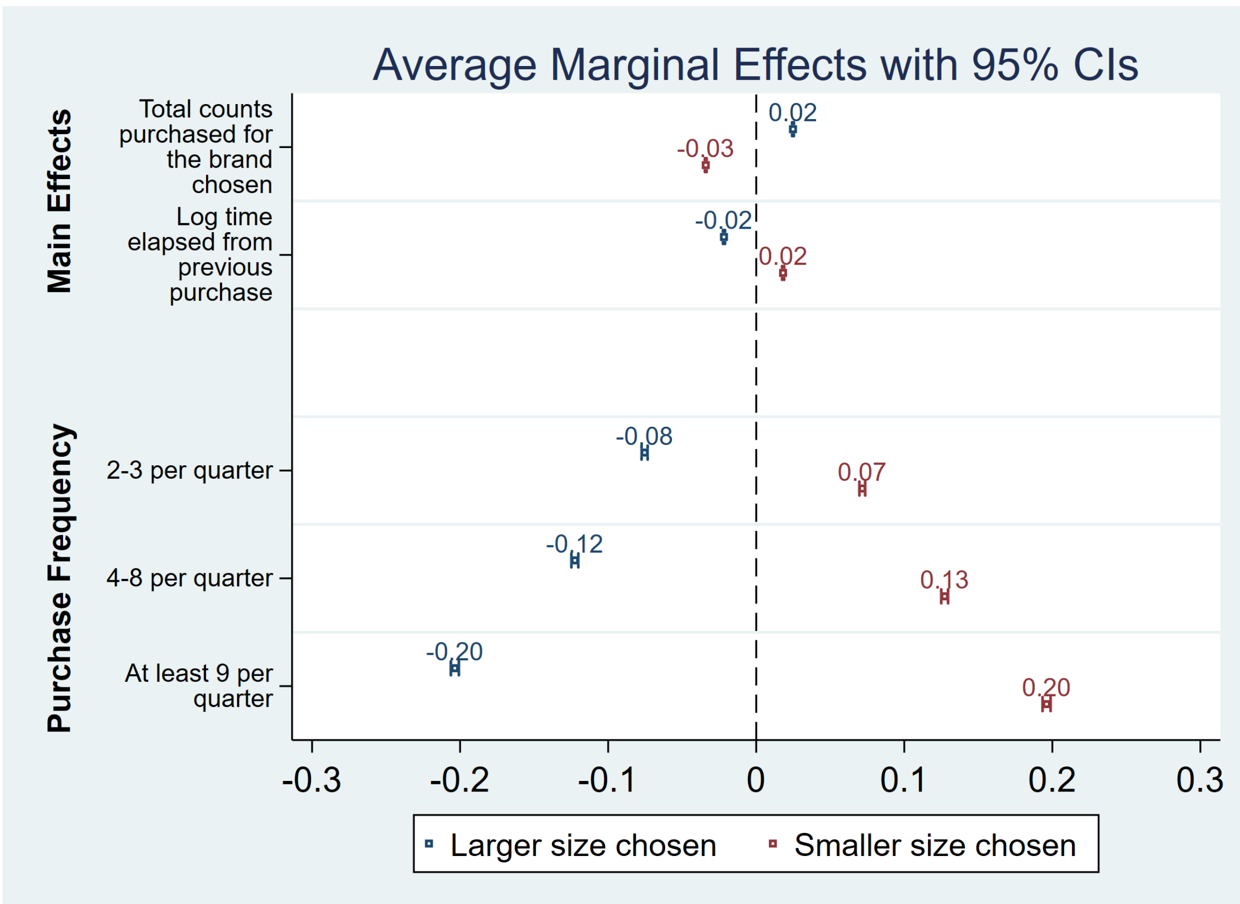



Figure 1.5: Reduced-Form Evidence in Yogurt Market: Average Marginal Effects of Initial Trial of a Brand on Size Choice

the consumption of the product. Before the trial of a brand, consumers have incomplete information on a product, but as they become more experienced, the uncertainty of the product decreases, and the probability of repurchase for the product increases. Furthermore, as uncertainty decreases, compared to their previous size choice, consumers have an incentive to choose a bigger size because there is a lower price per count with a bigger size purchase.

Compared to switching costs where it only depends on previous brand choice, since learning behavior results from the history of all purchases in each category, initial condition problems and decaying effects from each purchase should be considered to capture the learning behavior. While controlling for these two caveats, I test two competing theories: switching costs and learning behaviors. Then, I find three pieces of evidence: Rebuttal evidence on switching costs, suggestive evidence on learning from the fraction of brand switching cost over experience, and the average marginal effect of initial trials of a brand on size choice.

Rebuttal evidence on switching costs is that consumers' brand choice when consumers switched a brand was triggered by a temporary price cut in the previous shopping. According to the switching costs model, most consumers will not switch the brand again in their following shopping visit. However, only 45 percent of consumers repeat their purchases, and 36 percent of consumers switch to previous brands, again; whereas, 19 percent of consumers choose another brand. Learning behavior explains this behavior as persistent brand choice from learning or new trials on a brand.

Also, the fraction of brand switching over experience supports learning behavior and not switching costs. In switching costs model, the probability of switching costs is constant over purchases. However, the highest probability of brand switching is right after the first purchase of each brand; afterward, there is no statistical difference in the probability of brand switching over purchases. Since consumers learn the most information about the product during the initial trial of a brand, if the utility from a brand is less than the expected utility from the brand, they will switch to another brand during their next purchase. Additionally, the positive marginal effects of initial trials of a brand on smaller size choice for a multinomial logit model also supports learning behavior. Due to the consumer's uncertainty of product information, consumers choose to buy smaller sizes than what they used to buy when they first purchased a brand.

# Chapter 2

# EVIDENCE ON LEARNING BEHAVIOR WITH DOUBLE MACHINE LEARNING

## *2.1 Introduction*

In chapter 1, I investigate the association between the initial trial of a brand and the change of size choice. However, it still does not reveal the causal relationship. Therefore, to provide strong evidence on learning, in this chapter, with the same data used in Chapter 1, I adopt a double machine learning model, revisit the relationship between initial trial of a brand and smaller size choice, and show the causal channel of it.

The first argument is that experience to a brand updates information on true flavor or quality. This updated information reduces uncertainty. Hence, consumers are likely to purchase a larger size when they have complete information with the environment where the price per count at a larger size is lower than it is at a smaller size.

The basic problem in estimating the causal impact of learning on size-choice is that experience on a brand (Brand switching) is not randomly assigned, and it seems likely that there will be factors that are associated with both experiences on a brand and size choice. It is clear that any association between brand switching and the current size choice is likely to be spurious. The lack of random assignment makes establishing a causal link difficult without adequate controls. An obvious confounding factor is the existence of persistent individual-to-individual differences in location, such as urban or rural areas that are likely to be related to overall individual-level experience on a brand and size choice. For example, if consumers who live in rural areas visit the same store the only store accessible that supplies only one brand or one size and the store changes the brand, we could observe an association between initial trial of a brand and size choice.

On the other hand, it is also important to control flexibly for aggregate trends. For example, it could be the case that national trends of a brand were rising over some period while national trends of size were rising, but these trends were, instead, driven by entirely different factors. Without control for these trends, one would mistakenly associate an increase in size choice to the rise in experience on learning. In addition to these overall differences across individuals and times, there are other time-varying characteristics such as individual-level income, market characteristics, or the number of children that could be linked with experience on a brand and current size choice.

To address these confounds, following [24], I estimate a double machine learning model for individual-level size choice in which they condition on a number of these factors:

$$Y = D\theta_0 + g_0\left(X\right) + U, \quad E_P\left[U|X, D\right] = 0 \tag{2.1}$$

$$D = m_0\left(X\right) + V, \quad E_P\left[V|X\right] = 0 \tag{2.2}$$

where $E\left[U|X, D\right] = E\left[V|X\right] = 0$ for observation $i$ = 1,2, ..., n, $Y$ is log of growth rate of quantity purchased, $D$ is the initial trial on a brand, a vector of product characteristics and demographic characteristics, $X$, has a vector of $p$ controls, and $\theta_0$ measures the average causal or treatment effect of initial trial of a brand on changes in size. The nuisance parameters are $\eta_0 = (m_0, g_0)$.

With the main equation 2.1 and auxiliary equation 2.2, double machine learning estimates the causal effect $\eta_0$ as follows.

**Step1**: Obtain estimates of the conditional expectation functions $\hat{g}_0\left(X\right) = E[\hat{Y}|X]$ and $\hat{m}_0\left(X\right) = E[\hat{D}|X]$

**Step2**: Compute the residuals from main regression 2.1 and auxiliary regression 2.2 as $\hat{U} = Y - \hat{g}_0\left(X\right)$, $\hat{V} = D - \hat{m}_0\left(X\right)$ , respectively.

**Step3**: To estimate $\check{\theta}_0$, the causal effect, compute it as $\check{\theta}_0 = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i D_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i\left(Y_i - \hat{g}_0\left(X_i\right)\right)$

Compared to the econometrics approach and machine learning approach, double machine learning has several benefits in estimating the causal channels. First, working with high dimensional data-set, the double machine learning approach allows for non-parametric estimations and conducts variable selections. Thus, it allows for a broad set of covariates for identifying causal effects and reduces omitted variable bias and irrelevant variable selection. Although the econometrics approach specifies a set of controls by human judgment, which often omits important control variables or adds irrelevant variables. Conversely, the machine learning method could lead to regularization bias from bias-variance trade-off, and flexible machine learning models such as non-parametric estimations easily result in overfitting problems.

However, "debiasing" or "orthogonalization" in double machine learning alleviates regularization bias. Moreover, to remove bias from overfitting, double machine learning uses the following K-fold cross-fitting.

**Step1**. Split the sample into K equal parts, the auxiliary and main parts.

**Step2**. Use the auxiliary part to estimate the nuisance parameter and the main part to estimate the target parameter, obtaining one estimator of the target parameter.

**Step3**. By switching the roles of the main and auxiliary components, get another estimator of the target parameter.

**Step4**. Average the K estimators of the target parameter to obtain the final estimator.

### *2.2 Partially Linear IV Model*

My primary interest is to estimate the causal impact of a new trial brand on changes in size, holding the effect of the need to purchase a larger quantity constant.

In estimating the causal effect of a new trial brand on size choice, both omitted variable bias and reverse causality bias could be raised, in which I would need to resolve endogeneity problems by using instrumental variables. First, there could be omitted variables explaining size choice. For instance, when all family members show cough symptoms, regardless of the release of new cough drop products or changes in pricing, consumers will purchase a larger size of cough drops than they initially would have before. Since the family's need is usually not available in the data, the representative remedy is to obtain a proxy variable that is correlated to the omitted variable. By adding trend and week dummy variables, the model tries to remove omitted variable bias partially.

Moreover, reverse causality bias, where the dependent variable changes the independent variables could exist. For example, all family members have a cold, along with a sore throat. Suddenly, they need to purchase a larger quantity of cough drops. In this case, since consumers need a larger size of cough drops, they will choose the brand they experienced at the previous shopping trip. For this reason, the probability of purchasing a new brand will be small.

To overcome these biases, I need a source of exogenous variation in an initial trial of a brand. In this paper, I find that household income, and age and presence of children correlated with an initial trial of a brand, but not associated with size choice. Also, considering that as consumers have higher income, they are likely to show variety-seeking behavior, such as frequent changes in brand choice or purchase of multiple items. Besides, teenagers who are very curious to try new also are likely to try new brands. I also argue that income level, age, and presence of children is unlikely to influence a change of size choice. By this exclusion restriction, the above two variables are plausible instrument variables and capture the exogenous variation of an initial trial of a brand.

Consequently, I estimate the causal impact of a new trial brand on changes in size with a two-stage least-squares(2SLS) model while I isolate plausibly exogenous sources of variation in an initial trial on a brand. Then, based on a double machine learning partially linear model, I estimate the same impact and compare double machine learning estimates with the 2SLS estimate.

### *2.2.1 Specification*

In 2SLS estimator, I specify a linear function of all covariates, and the basic specification is

$$y_{it} = \theta_0 d_{it} + w_{it}'\beta + \delta_i + \gamma_t + \varepsilon_{it} \tag{2.3}$$

$$d_i = z_i'\Pi + x_i'\beta + \eta_i \quad E[\eta_i|x_i] = 0 \tag{2.4}$$

where $i$ indexes individuals, $t$ indexes time, $y_{it}$ is log of the ratio of quantity purchased at time t to the quantity purchased at time t-1, $w_{it}$ are a set of control variables to control for time-varying confounding individual-level factors $i$ at time $t$ including previous flavor choice, changes of flavor, time elapsed from the first purchase of each brand, log lagged total quantity consumed for each brand, change of store, deal, dummy indicating the second spell, previous brand choice, log relative price paid at the previous shipping, and log growth rate of price paid, $\delta_i$ are individual-specific effects that control any time-invariant individual-specific characteristics including dummy variables for regional market, $\gamma_t$ are time-specific effects that control flexibly for any aggregate trends including seasonality, and trend ($d_{it}$) is a dummy variable representing a new trial of a brand.

In a double machine learning model, I specified the following equation.

$$y_i = \theta_0 d_i + g_0\left(x_i'\beta\right) + u_i, \quad E[u_i|x_i, z_i'] = 0,$$

$$d_i = z_i'\Pi + m_0\left(x_i'\beta\right) + v_i \quad E[v_i|x_i] = 0$$

where $z_i$ is the instrumental variable, $y_i$ is log of the ratio of quantity purchased at time t to the quantity purchased at time t-1, $d_i$ is a dummy variable representing a new trial of a brand, and $x_i$ are a set of control variables, which includes time-varying confounding individual-level variables, individual specific variables, and time specific variables in the 2SLS estimator.

Since a double machine learning estimator applies to a large data set of control variables, the function $g_0$ is a highly predictive function of observables. It also estimates the coefficient on an endogenous variable in a partially linear instrumental variables model.

Estimations steps are as follows:

**Step1**: Partial out the effect of $x_i$ from $d_i$, $y_i$ and $z_i$ by machine learning methods.

**Step2**: Then we use the residuals to compute the IV estimator (2SLS) $\hat{\theta}$ of the parameter $\theta_0$.

**Step3**: Lastly, perform inference on $\theta_0$ using $\hat{\theta}$ and conventional heteroskedasticity robust standard errors.

## *2.3 Results*

### *2.3.1 OLS and IV Results*

Table 2.1 reports the OLS estimates and 2SLS estimates of interest, $\theta$ from equation 2.3, and Table 2.2 gives the corresponding first stages. In column (1) of Table 2.1, the OLS estimate shows the positive impact of a new trial of a brand on size change. If consumers purchase a larger size at time t than size purchased at time t-1, the ratio becomes greater than one, and the log form of the ratio is a positive number. Otherwise, the ratio becomes less than one and greater than zero, and the dependent variable becomes a negative number. Accordingly, it is interpreted for a positive coefficient of an initial trial on a brand that when consumers try a brand for the first time, they are likely to choose a larger size than their regular size choice from previous visits.

On the other hand, the 2SLS estimate of the impact of a new trial of a brand on change of size is highly significant with a standard error of 0.0113, and in fact, it has a negative coefficient, -0.0545. It is interpreted that consumers are likely to choose a smaller size than their original size choice in previous visits when they purchase a new brand. The difference between OLS estimate and 2SLS estimate suggests that reverse causality and omitted variable bias exist, and they are significantly large enough to change the conclusion.

Table 2.1: Regression Results of Changes in Size Choice

| Dependent Variable: | (1) Ordinary Least Squares | | (2) Two-Stage Least Squares | |
|---|---|---|---|---|
| Log (Growth rate of quantity choice + 1) | Coefficients | Standard Error | Coefficients | Standard Error |
| Initial trial on a brand | 0.0396*** | (0.0108) | -0.0545*** | (0.0113) |
| I($size_{t-1} < 25$) | 1.3197*** | (0.0198) | 1.9860*** | (0.0127) |
| I($49 < size_{t-1} < 99$) | 0.9942*** | (0.0167) | 1.5263*** | (0.0099) |
| I($98 < size_{t-1}$) | 0.5709*** | (0.0157) | 0.9280*** | (0.0092) |
| The ratio of coupon value to the tag price | -0.7682*** | (0.0228) | -0.7999*** | (0.0218) |
| Log ($date_t - date_{t-1}$) | -0.0207*** | (0.0017) | -0.0278*** | (0.0017) |
| I($Quarterly\ Shopping\ Frequency = 2$) | 0.0960*** | (0.0063) | 0.1371*** | (0.0070) |
| I($2 < Quarterly\ Shopping\ Frequency < 9$) | 0.0210*** | (0.0057) | 0.0440*** | (0.0066) |
| I($8 < Quarterly\ Shopping\ Frequency$) | 0.0241 | (0.0220) | -0.0132 | (0.0164) |

Notes:*p<0.1, **p<0.05, ***p<0.01. Standard errors in () of OLS regression are corrected for arbitrary correlation within households. Specifications of both regressions include the following variables: previous flavor choice, changes of flavor, time elapsed from the first purchase of each brand, log lagged total quantity consumed for each brand, change of store, deal, dummy for the second spell, previous brand choice, log relative price paid at the previous shipping, log growth rate of price paid, dummy for regional market, seasonality, and trend.

Table 2.2: First Stage for Initial Brand Choice

| Dependent Variable: Initial Brand Choice | First Stage for initial brand choice: Coefficient | Standard Error |
|---|---|---|
| Household income, compared to income $100,000+ | | |
| Under $5,000 | -0.3465 | (0.2255) |
| $5,000-$7,999 | -0.6448** | (0.3282) |
| $8,000-$9,999 | -0.8147*** | (0.2570) |
| $10,000-$11,999 | -0.0568 | (0.1948) |
| $12,000-$14,999 | -0.3733** | (0.1597) |
| $15,000-$19,999 | -0.2580** | (0.1268) |
| $20,000-$24,999 | -0.2200** | (0.1026) |
| $25,000-$29,999 | -0.1256 | (0.0971) |
| $30,000-$34,999 | -0.0641 | (0.0939) |
| $35,000-$39,999 | -0.0802 | (0.0966) |
| $40,000-$44,999 | -0.0832 | (0.1000) |
| $50,000-$59,999 | -0.1733* | (0.0989) |
| $60,000-$69,999 | -0.1659** | (0.0786) |
| $70,000-$99,999 | -0.1835** | (0.0880) |
| Age and Presence of Children, compared to under 6 & 13-17 | | |
| Under 6 | 1.2621*** | (0.4841) |
| 6-12 only | 1.1359** | (0.4618) |
| 13-17 only | 1.2129*** | (0.4538) |
| Under 6 & 6-12 | 1.5978*** | (0.4888) |
| 6-12 &13-17 | 1.3030*** | (0.4643) |
| Under 6 & 6-12 & 13-17 | 1.0352 | (0.6749) |
| No Children Under 18 | 1.2704*** | (0.4484) |

Notes: In the first stage of IV regression of changes in size choice, the following variables are included: male head's occupation and age, marital Status, discount rate, dummy variables for the second spell, lagged total quantity, change of store, deal, previous brand choice, log time elapsed from previous shopping visit, log lagged total quantity consumed for each brand, the time elapsed from the first purchase of each brand, previous flavor choice, changes of flavor, dummy variables for regional market, seasonality, and trend.

### *2.3.2 Double Machine Learning Results*

In Table 2.3, I report the DML estimate of ATE of an initial trial of a brand on size choice in a partially linear model. The considered ML methods and tuning parameters are the same as the previous examples except for the Ensemble method, from which we exclude Neural Network since the small sample size causes stability problems in training Neural Network. I consider 100 different sample partitions, use 5-fold cross-fitting to remove bias induced by overfitting, and present results using the median method. In brackets, I report standard errors adjusted for variability across the sample splits using the median method in brackets; and I report the median standard error from across the 100 splits in parentheses. First, the choice of the ML method used in estimating nuisance functions does not substantively change the conclusion in any of the examples, and I obtained broadly consistent results regardless of which method I employed. Second, the incorporation of uncertainty due to sample-splitting using the median method increases the standard errors relative to a baseline that does not account for this uncertainty, though these differences do not alter the main results in any of the examples. This lack of variation suggests that the parameter estimates are robust to the particular sample split used in the estimation of these examples.

In Table 18, I see uniformly large and negative point estimates across all procedures considered. Also, the estimated standard errors are small, so the estimated effects are statistically significant at the 1% level except for the estimated impact from the neural network model.

Even though the results differ from the baseline estimates reported in Table 2.1—an estimated coefficient of -0.05 with an estimated standard error of 0.01—they are qualitatively similar. Furthermore, both results indicate a strong and positive effect of a new purchase of a brand on a change of size choice.

The estimation results are consistent with the findings from the previous econometric regression model. The causal impact of a new brand trial on size choice is negative and significant across all estimation methods at the 5% level, regardless of the standard error

Table 2.3: Estimated Effect of Initial Brand Choice on Size Choice

| | Lasso | Trees | Boosting | Forest | Nnet | Ensemble | Best |
|---|---|---|---|---|---|---|---|
| Median ATE | -0.35 | -0.45 | -0.37 | -0.46 | -0.02 | -0.45 | -0.45 |
| SE (median) | [0.11] | [0.11] | [0.13] | [0.11] | [0.43] | [0.11] | [0.11] |
| SE | (0.03) | (0.09) | (0.1) | (0.1) | (0.02) | (0.1) | (0.1) |

Notes: Estimated ATE and standard errors from a linear model is based on orthogonal estimating equations. Column labels denote the method used to estimate nuisance functions. Results are based on 100 splits with point estimates calculated the median method. The median standard errors across the splits are reported in parentheses, and standard errors calculated using the median method to adjust for variation across splits are provided in brackets.

estimator used.

## *2.4 Conclusion*

Many economists and social scientists in the field of economics and marketing have agreed on consumers' dynamic choice and state dependence. However, even though theories such as learning and switching costs exist, sources of state-dependent choice are still debatable because we are lack empirical evidence derived from unobservable learning and switching costs behavior.

In this study, I test two competing theories based on the implication of learning behaviors and switching costs behaviors and find strong evidence of learning behavior. Especially, I find the negative impact of a new purchase of a brand on change of size, and it shows that consumers respond to the uncertainty of information on a brand. Compared to the previous evidence of learning from descriptive analysis, the reduced form of evidence in Chapter 1 and Chapter 2 convince the learning behavior as a source of state-dependence for consumers who shows persistence in brand choice.

Nevertheless, learning behavior theory cannot explain the state dependence choice fully in that learning behavior only explain persistent brand choice, one of the state dependence choice, and fail to explain why consumers switch to other brands. For example, when consumers are fully informed about all brands in the market, both switching costs and learning behavior does not explain brand switching decision. If the brand switching is also a state-dependent choice, we need other explanations to understand consumers' dynamic choice. According to [62, 52], satiation behavior that describing diminishing marginal utility over quantity purchased may explain why consumers switch brands. It also represents state dependence where past choice directly influences the present decision. As consumers' previous quantity choice of a brand decreases current utility from the brand, consumers do not purchase the brand and switch to other brands.

Therefore, in Chapter3, I develop a dynamic consumer panel demand model incorporating both decisions: persistent brand choice and brand switching. Also, I test whether consumers show learning behavior or switching costs behavior by allowing for brand switching decision.

# Chapter 3

# DYNAMIC PANEL DEMAND MODEL INCORPORATING VARIETY-SEEKING BEHAVIOR

## *3.1 Introduction*

In this chapter, I propose a structural dynamic panel model incorporating variety-seeking consumers who purchase multiple items at a time or frequently switch products, based on a multiple-discrete choice model. The reason why I build this model is that previous studies only focus on learning behavior or switching costs behavior and not satiation in variety-seeking behavior: a diminishing marginal utility over quantity consumed. Most studies do not focus on satiation in variety-seeking behavior despite that variety-seekers' choices also depend on past decisions and satiation in variety-seeking behavior is not considered as sources of dynamic choice. Furthermore, to understand how preferences for brands evolve over experience, the time-varying effects of state-dependence also needs to be incorporated.

Hence, I specify a dynamic panel model that tests the two competing theories by allowing brand preference parameters to vary by past choices and time elapsed from each purchase. Then, I estimate utility parameters at consumer levels by applying a Bayesian Markov Chain Monte Carlo method, which uses the Metropolis-Hasting(M-H) algorithm. This estimation infers the posterior distribution of parameters for no-purchase of a brand, one-time purchase of a brand, two-time purchases of a brand, and so forth.

With a rich data set from a Nielsen consumer panel that tracks the majority of shopping trips for cough drops by 28.724 households in the US between 2006 and 2015, I find that the primary source of dynamic choice is learning after controlling the differences in preferences across consumers. However, only the first purchase of a brand affects current brand choice, which means that the consumer learns fast. This result is also contrary to a slowdown learning model, such as a Bayesian learning model. Also, the model incorporating satiation behavior explains data better than the model ignoring it.

The rest of the chapter proceeds as follows. Section 2 describes data on consumers' purchase patterns supporting multiple-discreteness choice. Section 3 presents a dynamic panel demand model integrating state-dependence and purchase of multiple items based on the multiple-discrete choice models. Then, I test whether brand preferences estimate are consistent with (i) the switching costs model, (ii) the Bayesian learning model, (iii) satiation in variety seeking, (iv) switching costs with satiation in variety seeking, or (v) non-Bayesian learning with satiation in variety seeking with the dynamic panel demand model. By tracking how preferences for brands have changed over experiences, I find the consistent results with reduced-form evidence that consumers learn fast in Section 4. Section 5 provides robustness checks. I conclude the paper with a summary in Section 6.

### *3.2 Data: Multiple-Discrete Choice*

To check the extent of multiple discrete choices in the transaction data, I compute the number of purchases with multiple items and the number of purchases with only one item. In the sample, 30% of the transactions switch brands. 3% of the transactions involve a purchase of both brands. 7.7% of the transactions involve a purchase of various flavors.

Table 1 displays information on the frequency of both single item purchases and multiple item purchases. 92% of the transactions were corner solutions where only one of the nine varieties was selected, and 8% of the transactions were interior solutions where more than one variety was purchased. This simultaneous purchase of multiple varieties requires me to use a multiple-discrete choice model, not a standard discrete choice model. The data indicates that there is no dominant product offering or characteristic in Table 1. A portion of offerings in Table 1 also indicates that consumers' preferences are related to the characteristics of the offerings. By using multiple-discrete choice models, this variety-seeking behavior is incorporated. It is noted that these variety-seekers who purchased multiple items are not learners who experiment with multiple brands. This is because the fraction of variety-seeking

Table 3.1: Purchase Summary

| Variety | Purchase incidence | Mean of total purchase quantity (std.dev) | Single item purchase | Multiple items purchase |
|---|---|---|---|---|
| Halls1 | 526 | 50.18 (41.12) | 474 | 52 (9.9%) |
| Halls2 | 429 | 48.75 (31.75) | 389 | 40 (9.3%) |
| Halls3 | 341 | 55.50 (45.47) | 314 | 27 (8%) |
| Halls4 | 148 | 31.99 (14.70) | 138 | 10 (6.8%) |
| Halls5 | 189 | 53.46 (33.30) | 169 | 20 (10.6%) |
| Halls6 | 101 | 37.18 (20.86) | 90 | 11 (10.9%) |
| Halls7 | 93 | 64.09 (67.45) | 93 | 0 (0%) |
| Ricola1 | 267 | 31.97 (19.80) | 260 | 7 (2.6%) |
| Ricola2 | 93 | 24.92 (14.55) | 85 | 8 (8.6%) |
| Total | 2,187 | 48.44 (39.70) | 2,012 (92%) | 175 (8%) |

behavior is constant over shopping trips, and learners purchase various products only in the initial shopping trips.

Table 2 shows the distinct characteristics of selected brands. The price of Ricola per count is 0.08-0.102, which is almost twice as high as the price of Halls per count, 0.035-0.064. Additionally, the most and second most frequently purchased flavor of Halls are menthol and eucalyptus flavors. The most frequently purchased flavor of Ricola is herb flavor, which is similar to, but not the same as Halls' flavors. Also, a consumer who has a strong preference for sugar-free is likely to choose Halls opposed to Ricola.[1] Therefore, consumers' changes in preference of flavors or sugar-free characteristics could lead to brand switching. To avoid confounding structural state dependence with unobserved preference differences, a Bayesian

[1]Results on regression of brand choice are available from the author upon request.

Table 3.2: Product Varieties and Characteristics

| Variety(j) | Obs. | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Purchase incidence | Halls | Ricola | Eucalyptus | Herb | Cherry | Menthol | Honeylemon | Sugar free | Throat | Price per count |
| Halls1 | 526 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0.043 (0.017) |
| Halls2 | 429 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0.044 (0.018) |
| Halls3 | 341 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0.047 (0.020) |
| Halls4 | 148 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0.067 (0.022) |
| Halls5 | 189 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0.035 (0.015) |
| Halls6 | 101 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0.063 (0.018) |
| Halls7 | 93 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.064 (0.021) |
| Ricola1 | 267 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.080 (0.023) |
| Ricola2 | 93 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0.102 (0.036) |

model of heterogeneity is specified in the main model.

### 3.3 Structural Model

In this section, I propose a dynamic panel demand model that can accommodate state-dependence and a purchase of multiple items. Following the multiple-discrete choice models of [61] and [52], I assume that a consumer's utility is a sum of utility from each product purchased. I define consumer $i$'s utility over product $j = 1, ..., J$ at shopping trip $t$ as:

$$U_{it} = \sum_{j=1}^{J} \psi_{ijt} \left(q_{ijt} + 1\right)^{\alpha_{ijt}}, \tag{3.1}$$

$$MU_{it} = \frac{\partial U_{it}}{\partial q_{ijt}} = \alpha_{ijt}\psi_{ijt} \left(q_{ijt} + 1\right)^{\alpha_{ijt}-1},$$

where $q_{ijt}$ is the quantity chosen for product $j$ by consumer $i$ at shopping trip $t$, and demand parameters consist of the baseline level of utility, $\psi_{ijt}$, and satiation parameter $\alpha_{ijt}$ that affects the rate at which marginal utility diminishes. Equation (1) defines a valid utility function under the restrictions that $\psi_{ijt} > 0$ and $0 < \alpha_{ijt} \leq 1$. If a particular variety has a high value of $\psi_{ijt}$ and a value of $\alpha_{ijt}$ close to one, then we would expect to see purchases of large quantities of only one variety (high baseline preference and low satiation). On the other hand, small values of $\alpha_{ijt}$ imply a high satiation rate and I expect to see multiple varieties purchased if their values of $\psi_{ijt}$ are not too different. The parameter is given by $\frac{\alpha_{ijt}}{1-\alpha_{ijt}} = exp\left(\alpha^*_{ijt}\right)$ and $\psi_{ijt} = exp(X_j^{'}\beta_{it})$ where $X_j$ is a K-dimensional row-vector of product j's characteristics and errors $\varepsilon_{ijt}$. The utility specification in Equation (1) can accommodate a wide variety of situations, including purchases involving a large number of different varieties as well as purchases with only one variety selected.

To identify structural state dependence, I specify that demand parameters, $\beta_{it}$ and $\alpha_{it}$, depend on past choices and the time elapsed from each purchase:

$$\begin{aligned} \beta_{it} &= \theta_i^{'} h_{it} + e_{it}, \quad e_{it} \sim N\left(0, V_\beta\right), \\ \alpha_{ijt} &= \theta_{ij}^{\alpha\,'} h_{it} + e_{ijt}^{\alpha}, \quad e_{ijt}^{\alpha} \sim N\left(0, V_\alpha\right). \end{aligned} \tag{3.2}$$

where $h_{it}$ is a vector of the history of past choices, and the time elapsed from each purchase.[2] Standard demand model specifications ignore dynamics in (1), where error terms were assumed to be i.i.d., but they could be serially correlated owing to state-dependence or unobserved heterogeneity.[3] To show structural state dependence, I specify the demand parameters depend on the history of past choices, $h_{it}$. I first test the statistical significance of my estimates of state dependence, $\theta$, in equation (2) while controlling for consumer heterogeneity with a Bayesian approach. Next, I investigate behavior mechanisms that generate dynamics, with the following five cases: switching costs, Bayesian learning, variety-seeking, switching costs with variety-seeking, and Non-Bayesian learning with variety-seeking.

### *3.3.1 Switching Costs*

The key difference of specification between switching costs and learning model is whether previous brand choice summarizes the history of past purchases or each brand choice in the past matters. According to the switching costs model, switching costs incur if the consumer chooses any other brands except for the previous brand chosen, and it is interpreted as a psychological switching cost or a utility cost. Therefore, with a persistent brand choice, current utility includes the benefit of not paying switching costs. In the specification, coefficient on a lagged choice, $\theta^{lag}_{ik,1}$, captures switching costs and it is also called the state dependence coefficient.[4] Switching costs are defined similar to [33] and [42] as constant over time consumer-specific costs, $\theta^{lag}_{ik,1}$, known to the consumers. Since satiation behavior is ignored in this model, I set the satiation parameter as one, $\alpha_j = 1$. The baseline parameter

[2] As [82] pointed out, the effects of state dependence vary over time, and this change could be owing to the decaying effect of state dependence which accumulated at each purchase. Also, the degree of decaying could differ at each purchase. For example, in Bayesian learning, the degree of decaying for the information acquired at the first purchase and that of at the fourth purchase could be different.

[3] Previous literature states that not controlling for unobserved consumer differences or marketing-mix variables could lead to auto-correlated errors. See, for example, [53], [60], [82], and [81]. To disentangle state-dependence from unobserved heterogeneity, literature exploits the difference between stated preference and revealed preference, choice-consumption mismatch strategy, and variation in observable product characteristics across markets.[85, 90, 83]

[4] See, for example, [82],[76], [33].

$\beta_{it,k}$ is specified as associated with lagged choices, and the time elapsed from the previous shopping date of the brand of product $j$:

$$\begin{aligned}\beta_{it,k} &= \theta_{ik}^{'} h_{it} + e_{it,k} \\ &= \theta_{ik,0}^{lag} + \theta_{ik,1}^{lag} I(d_{ij,t} = d_{ij,t-1}) + \theta_{ik,2}^{lag} (t - t_{lag}) I(d_{ij,t} = d_{ij,t-1}) + e_{it,k},\end{aligned}$$

where $d_{ij,t}$ =1 if household $i$ purchases the brand of product $j$ at shopping trip $t$, so if household $i$ does not change their brand choice compared to the choice at shopping trip *t-1*, $I(d_{ij,t} = d_{ij,t-1}) = 1$. Therefore, $\theta_{i1}^{lag}$ is interpreted as switching costs indicating how much utility they will lose when they switch to other brands. Also, I control for decaying effect with $t - t_{lag}$, the time elapsed from the previous purchase of brand $j$ to shopping trip $t$. The estimates of the decaying effect, $\theta_{ik,2}^{lag}$, is also interpreted as how much the estimates $\beta_{it}$ change when one day is elapsed from the previous purchase of the brand of product $j$.

### *3.3.2 Bayesian Learning*

According to a Bayesian learning model, consumers have imperfect information on unobserved characteristics of a product. They update information on unobserved characteristics of product $j$ at every purchase, and this information leads to state-dependent choice [29]. Thus, not only the lagged choice but also the consumer's entire history of past choices contribute to updating information and forming state dependence. Also, if a diminishing marginal return of learning exists, brand choices in the initial shopping periods are more likely to affect present brand choice rather than a lagged brand choice. Therefore, I added dummy variables representing how many times the consumer experienced the brand, $d_{ijt_1}, d_{ijt_2}, ..., d_{ijt_n}$, and the coefficient of each of the dummy variables shows the return of learning from each purchase.

Suppose that consumer i chooses product $j$ and she has purchased the brand of product $j$ $n$ times before shopping trip t. Here, I set the satiation parameter as $\alpha_j = 1$ in the switching costs model. The baseline parameter $\beta_{it,k}$ is specified as associated with all past choices and the time elapsed from each purchase of the brand of product j:

$$\beta_{it,k} = \theta_{ik,0} + \theta_{i1} d_{ijt_1} + \cdots + \theta_{ik,n} d_{ijt_n} + \theta_{ik,n+1} \left(t - t_{j1}\right) d_{ijt_1} + \cdots + \theta_{ik,2n} \left(t - t_{jn}\right) d_{ijt_n} + e_{it,k},$$

where $I(d_{ijt_n}) = 1$ when consumer $i$ already experienced the brand of product $j$ $n$ times at shopping trip $t$, the estimates $\theta_{ik,n}$ indicates the information accumulated from $n$ times purchases of the brand of product $j$. On the other hand, if $n = 0$, since the consumer does not experience the brand of product $j$ at all, $\theta_{ik,0}$ shows priors for the brand of product $j$ for *k=1,2* and preference for product characteristics *k=3,...,9*. It should be noted that $\theta_i$ is a preference vector of K product characteristics. Also, decaying effects matter because the updated information at every purchase will be decayed. $(t - t_{jn})$ is a variable of the time elapsed from the $n$ th purchase of the brand of product $j$ by consumer $i$ to shopping trip $t$.[5]

If consumers show both learning behavior and persistent brand choice, the estimates of $\theta_i$ should increase over experiences as in Equation (3).

$$\theta^{all}_{ik,0} < \theta^{all}_{ik,1} < \theta^{all}_{ik,2} < \theta^{all}_{ik,3}, \quad k = 1, 2. \tag{3.3}$$

### *3.3.3 Satiation in Variety Seeking*

Under the general multiple-discrete choice model in Equations (1) and (2), variety-seeking behavior is specified with satiation parameters and baseline utility parameters as follows:

$$\alpha_j \sim N\left(\bar{\alpha}, V_\alpha\right), \quad \beta_i \sim N\left(\bar{\beta}, V\right).$$

### *3.3.4 Switching Costs with Satiation in Variety Seeking*

Under the above multiple-discrete choice model, switching costs are incorporated in the specification of satiation parameters and baseline utility parameters as:

---

[5] In this model, I set the maximum number of purchases as 5, $n = 1, ..., 5$. In a separate appendix available upon request, I show the result of a multinomial random effect model including brand intercept and dummy indicating the total number of purchases for brands. I find that marginal effect does not increase after three purchases of a brand. Therefore, it is enough for this specification to set the maximum number of purchases as 5.

$$\alpha_{ijt} = \theta^{\alpha}_{i,j} + \theta^{\alpha}_{i,0} + \theta^{\alpha}_{i,1} I(d_{ij,t} = d_{ij,t-1}) + \theta^{\alpha}_{i,2}\,(t - t_{lag})\, I(d_{ij,t} = d_{ij,t-1}) + e^{\alpha}_{ijt},$$
$$\beta_{it,k} = \theta^{lag}_{ik,0} + \theta^{lag}_{ik,1} I(d_{ij,t} = d_{ij,t-1}) + \theta^{lag}_{ik,2}\,(t - t_{lag})\, I(d_{ij,t} = d_{ij,t-1}) + e_{it,k},$$

where both the satiation parameter and the baseline parameter are heterogeneous, and they depend on the lagged choices and the time elapsed from the previous purchase date of the brand of product $j$.

### *3.3.5 Non-Bayesian Learning with Satiation in Variety Seeking*

Under the above multiple-discrete choice model, non-Bayesian learning is incorporated in the specification of the baseline utility parameter and the satiation parameter as:

$$\begin{aligned}
\alpha_{ijt} =& \theta^{\alpha}_{i,j} + \theta^{\alpha}_{i,0} d_{ijt_0} + \theta^{\alpha}_{i,1} d_{ijt_1} + \cdots + \theta^{\alpha}_{i,n} d_{ijt_n} \\
&+ \theta^{\alpha}_{i,n+1}\,(t - t_{j1})\, d_{ijt_1} + \cdots + \theta^{\alpha}_{i,2n}\,(t - t_{jn})\, d_{ijt_n} + e_{it}, \\
\beta_{it,k} =& \theta_{ik,0} + \theta_{i,1} d_{ijt_1} + \cdots + \theta_{ik,n} d_{ijt_n} \\
&+ \theta_{ik,n+1}\,(t - t_{j1})\, d_{ijt_1} + \cdots + \theta_{ik,2n}\,(t - t_{jn})\, d_{ijt_n} + e_{it,k},
\end{aligned}$$

where $d_{ijt_n}$ is the indicator function which equals one if consumer $i$ purchased the brand of product $j$ $n$ times at shopping trip $t$, and $(t - t_{jn})$ presents the time elapsed from the $n$th purchase of the brand of product $j$ by consumer $i$ at shopping trip $t$ for $n = 1, ..., 5$.

Compared to previous learning models, the non-Bayesian learning model captures diminishing returns to experience from the satiation parameter. Thus, the satiation parameter explains both simultaneous demands for variety and sequential order for variety because a satiation parameter is specified as a history of past choices.

I find that any of the models 1,2,4,5 incorporating dynamics into the baseline parameters $\{\psi_{ijt}\}$ leads to a dramatic improvement in model fit. Also, a model with a heterogeneous parameter for the satiation parameter specified with the brand choice history is preferred to a model with a common parameter for the brand of product $j$. Appendix Table A7 reports the results of the regression on the optimal quantity. The optimal quantity is derived from consumers' utility maximization problem. The results support the specification of Model 5 where the satiation parameter depends on the history of past choices.

## 3.4 Estimation

Here, I discuss the details of the estimation for the multiple-discrete choice model. As I only observe each consumer's shopping trip, $t$ represents the order of shopping trips, not calendar time. Also, since the total number of observations for each consumer varies, the panel data-set is an unbalanced panel. My statistical specification consists of (1) the within unit likelihood function and (2) the across time and unit variation in parameters which is called "unobserved heterogeneity". This joint model is estimated using a hybrid Markov Chain Monte Carlo (MCMC) approach. This approach allows me to infer individual-level parameters so that I can test sources of dynamics in brand choice. For example, See [51].

To develop a statistical specification, I follow a standard random utility approach and introduce a multiplicative normal error into marginal utility:

$$ln(U_{ijt}) = ln(MU_{ijt}) + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim N(0,1),$$

where $MU_{ijt}$ is the derivative of the utility function in Equation (1) with respect to $q_{ijt}$. I use a log-normal error term to enforce the positivity of marginal utility. I assume that $\varepsilon_{ijt}$ has an identity covariance structure for simplicity. In my empirical analysis, I consider the demand for different cough drop flavors and do not feel that this assumption is too restrictive. Also, it is assumed that the consumer knows the value of $\varepsilon_{ijt}$, but is not observable to the econometrician.

### *3.4.1 Within Unit Likelihood*

To derive the likelihood and the optimal demand, $q^*_{ijt}$, I use the Kuhn-Tucker conditions from the utility maximization problem to solve for the optimal demand. I obtain the likelihood function by maximizing Equation (1) subject to a budget constraint($\sum_{j=1}^{J} p'_{ijt} q_{ijt} \leq E_{it}$). By differentiating the Lagrangian $L = U_{ijt}(q) + \lambda \left( E_{it} - \sum_{j=1}^{J} p'_{ijt} q_{ijt} \right)$, it gives me the standard Kuhn-Tucker first-order conditions as follows:

$$\frac{\partial L}{\partial q_{ijt}} = \alpha_{ijt} \psi_{ijt} (q_{ijt} + 1)^{\alpha_{ijt} - 1} e^{\varepsilon_{ijt}} - \lambda p_{ijt} = 0, \quad q^*_{ijt} > 0, \tag{3.4}$$

$$\frac{\partial L}{\partial q_{ijt}} = \alpha_{ijt} \psi_{ijt} (q_{ijt} + 1)^{\alpha_{ijt} - 1} e^{\varepsilon_{ijt}} - \lambda p_{ijt} < 0, \quad q^*_{ijt} = 0, \tag{3.5}$$

where $p_{ijt}$ is the price of product $j$ consumer $i$ paid at shopping trip $t$ , $E_{it}$ is the total expenditure of consumer $i$ at shopping trip $t$, $\lambda$ is the Lagrange multiplier, and $q^*_{ijt}$ denotes the optimal quantity demanded for each of the $J$ goods under consideration.

Dividing by price and taking logs, the Kuhn-Tucker conditions (5) and (6) can be rewritten as:

$$\varepsilon_{ijt} = -ln \left( \alpha_{ijt} \psi_{ijt} (q_{ijt} + 1)^{\alpha_{ijt} - 1} \right) + ln (p_{ijt}) + \varepsilon_{i0t}, \quad if\ q^*_{ijt} > 0, \tag{3.6}$$

$$\varepsilon_{ijt} < -ln \left( \alpha_{ijt} \psi_{ijt} (q_{ijt} + 1)^{\alpha_{ijt} - 1} \right) + ln (p_{ijt}) + \varepsilon_{i0t}, \quad if\ q^*_{ijt} = 0. \tag{3.7}$$

Setting $v_{ijt} = \varepsilon_{ijt} - \varepsilon_{i0t}$ and $h_{ijt}(q*, p) = -ln\alpha_{ijt} - X'_j \beta_{it} - (\alpha_{ijt} - 1) ln (q_{ijt} + 1) + ln (p_{ijt})$, for $j = 1, .., J$, I define the likelihood function as (temporarily dropping the subscripts i and t to improve readability):

$$\begin{aligned} & Pr \left( q^*_j > 0\ and\ q^*_l = 0; j = 1, ..., n\ and\ l = n+1, ..., J \right) \\ & = \int_{-\infty}^{h_j} \cdots \int_{-\infty}^{h_{n+1}} \varphi (h_1, ..., h_n, v_{n+1}, ..., v_J | \Omega) |J| \, dv_{n+1} \cdots dv_J, \end{aligned} \tag{3.8}$$

where $\varphi(\cdot)$ is the normal density, $h_j = h_j(q^*, p)$, $v = (v_2, ..., v_m)' \sim N(0, \Omega)$ where $\Omega$ is the covariance matrix of the differenced errors with element $v_j$, and $J$ is the Jacobian of the transformation from $\{v_j\}$ to $\{q^*\}$, which is with elements given by,

$$J_{jl} = \frac{\partial h_{j+1}(q^*, p)}{\partial q^*_{l+1}}, \quad j, l = 1, ..., n-1.$$

Thus, the likelihood function has a density component corresponding to the goods with nonzero quantities for the first $n$ of $J$ alternatives, $\varphi(h_1, ..., h_n | \Omega) |J|$, and a mass function corresponding to the corners in which some of the goods will have zero optimal demand, $\int_{-\infty}^{h_J} \cdots \int_{-\infty}^{h_{n+1}} \varphi\left(v_{n+1}, ..., v_J\right) \, dv_{n+1} \cdots dv_J$.

I use the Geweke Hajivassiliou Keane (GHK) simulator to compute the multivariate normal integral in (9). (See [59], [49], and [13] for the case of extreme value errors.)

### *3.4.2 Heterogeneity, Prior Distributions, and Identification.*

I allow for heterogeneous parameters by specifying random-effects distribution for model coefficient vector. The baseline parameters related product characteristics are heterogeneous, $\beta_{it} = \theta_i^{'} h_{it} + e_{it}$. The satiation parameter only relating to brands are also heterogeneous, $\alpha_{ijt} = \theta_{ij}^{\alpha}{}' h_{it} + e_{ijt}^{\alpha}$.

$$\beta_{it} \sim N\left(\bar{\theta}' h_{it}, V_\beta\right), \alpha_{ijt} \sim N\left(\bar{\theta}_j^{\alpha\prime} h_{it}, V_\alpha\right), \tag{3.9}$$

$$\theta_i \sim N\left(\bar{\theta}, V_\theta\right), \theta_{ij}^{\alpha} \sim N\left(\bar{\theta}_j^{\alpha}, V_{\theta^\alpha}\right). \tag{3.10}$$

where $i = 1, ..., N$ indexes the households; $t = 1, ..., T_i$ indexes the shopping trip; and to identify state-dependence, all demand parameters are specified as associated with the history of past choices, $h_{it}$, and the estimated matrix of coefficients on all past choices $\theta_{ij}$ in equation (2) shows whether state-dependence exists or auto-correlated taste shocks exist. I used a Bayesian procedure to estimate the parameters in the model and specify a hierarchical prior with a normal distribution as the first-stage priors.[6] Prior distributions and the Markov Chain Monte Carlo (MCMC) estimation algorithm appear in Appendix B. I refer the reader to Rossi, Allenby, and McCulloch (2005) [77] for a more detailed discussion. Estimation was based on 20,000 MCMC iterations with the first 10,000 as burn-in.

---

[6]The hierarchical prior provides a convenient way of specifying an informative prior which avoids the problem of overfitting.

## 3.5 Empirical Application

### 3.5.1 Model Evaluation and Comparison

I compare 5 models that can be classified into three broad groups: (1) a switching model with no satiation, but heterogeneous baseline parameters depending on lagged brand choice and time elapsed from the previous choice, (2) a learning model with no satiation, but heterogeneous baseline parameters depending on all past choice and time elapsed from each choice, and (3) multiple-discreteness models with satiation parameter. All five models allow for heterogeneous baseline parameters. However, models 1-3 allow for homogeneous satiation parameters across households, while models 4-5 allow for heterogeneous satiation parameters. Also, demand parameters in model 1 and 4 are specified as a lagged brand choice, but these of model 2 and 5 are specified as all past brand choices. Previous literature covers only models in (1) and (2), and these models provide a comprehensive set for assessing the structural state dependence and a source of state dependence. The two measures of the fitting model, log marginal density and the Bayesian deviance information criterion (DIC) for model 1-5, are reported in Table 5. The DIC is a measure of model comparison that explicitly penalizes a model for its number of parameters ([89]). Model 3 provides the best fit compared to Model 1 and Model 2. This means that specifying satiation parameters improves fit dramatically. Therefore, both state dependence parameters and satiation parameters are specified in Model 4 and Model 5. However, the fit of Model 4 and Model 5 is worse than Models 1-3. The main reason is that I specify that demand parameters relating to product characteristics depend on past choices to distinguish between the effect of state dependence and the effect of reinforced preference on flavors, even though coefficients on product characteristics do not depend on past choices. According to the result of regression on quantity choice, satiation parameters depend on past choices, and coefficients of past choices are statistically significant. Hence, this supports the specification of model 5. Also, both model 4 and model 5 capture the effect of learning. The difference between model 4 and model 5 is that the coefficient on lagged choice in model 4 shows the benefit of information accumulated from all purchases on the brand chosen in the last shopping trip; although, the coefficients on each past choice in model 5 show the benefit of information at each purchase. Therefore, to show the learning process, I report the results of model 5.

### *3.5.2 Parameter Estimates*

I employ Bayesian MCMC methods to evaluate the posterior joint density for these models. Appendix B lists the algorithms for model estimation. All models were estimated using 20,000 iterations of the Markov Chain. For all models, I used the last 10,000 draws to obtain parameter estimates.

Parameter estimates for Model 5 are reported in Table 6. The second column of Table 6 indicates estimates of satiation parameter, $\alpha$. In general, the parameters are estimated precisely, and nearly all have substantial posterior mass away from zero.

There is considerable evidence of state dependence on brand choice and satiation behavior. A value of $\alpha$ close to one indicates less curvature in the utility function, implying less satiation, while a value of $\alpha$ close to zero indicates greater satiation. As satiation increases, optimal demand (x*) for each variety becomes smaller with the purchase of multiple varieties, holding all else constant. In Table 7, the mean of estimates of satiation parameters for Halls's products is from 0.494 to 0.526. On the other hand, for Ricola products, it is from 0.527 to 540. This indicates that consumers are more satiated with Halls's products than Ricola products do. Estimates of the mean of the random effects distribution for $\beta$ are positive in the Table 7.[7] The most favorable characteristic, on average, is C9 with a coefficient of 0.54, and the least favorable characteristic is C1 with a coefficient of 0.49. Heterogeneity around these mean levels is large, with some households favoring and some households disliking each of the characteristics. The most diverse preferences are for characteristic C1 with a variance of 0.028. Also, there are a number of large negative correlations, with respondents favoring

[7] A positive element indicates that the majority of respondents in the survey prefer to have more of the characteristics than less, while a negative coefficient indicates that a majority would prefer to have less of the characteristic.

Table 3.3: Model Comparison

| Model | (Satiation, Baseline) | Log-marginal Likelihood* | DIC |
|---|---|---|---|
| 1. Switching costs model | $\alpha = 1$<br>$\beta_{it,k} = \theta^{lag}_{ik,0} + \theta^{lag}_{ik,1} I(d_{ij,t} = d_{ij,t-1}) + \theta^{lag}_{ik,2}(t - t_{lag}) I(d_{ij,t} = d_{ij,t-1}) + e_{it,k}$ | -8921 | 35659 |
| 2. Bayesian learning model | $\alpha = 1$<br>$\beta_{it,K} = \theta_{ik,0} d_{ijt_0} + \theta_{ik,1} d_{ijt_1} + \cdots + \theta_{ik,n} d_{ijt_n} + \theta_{ik,n+1}(t - t_{j1}) d_{ijt_1} + \cdots + \theta_{ik,2n}(t - t_{jn}) d_{ijt_n} + e_{it,k}$ | -9012 | 18028 |
| 3. Variety model | $\alpha_j \sim N(\bar{\alpha}, \sum_\alpha)$<br>$\beta_i \sim N\left(\bar{\beta}, \sum_\beta\right)$ | -4050 | -19055 |
| 4. Switching costs with satiation in variety model | $\alpha_{ijt} = \theta^{\alpha}_{i,j} + \theta^{\alpha}_{i,0} + \theta^{\alpha}_{i,1} I(d_{ij,t} = d_{ij,t-1}) + \theta^{\alpha}_{i,2}(t - t_{lag}) I(d_{ij,t} = d_{ij,t-1}) + \theta^{\alpha}_{i,3} + e^{\alpha}_{ijt}$<br>$\beta_{it,k} = \theta^{lag}_{ik,0} + \theta^{lag}_{ik,1} I(d_{ij,t} = d_{ij,t-1}) + \theta^{lag}_{ik,2}(t - t_{lag}) I(d_{ij,t} = d_{ij,t-1}) + e_{it,k}$ | -10265 | 16157 |
| 5. Non-Bayesian learning with satiation in variety model | $\alpha_{ijt} = \theta^{\alpha}_{i,j} + \theta^{\alpha}_{i,0} d_{ijt_0} + \theta^{\alpha}_{i,1} d_{ijt_1} + \cdots + \theta^{\alpha}_{i,n} d_{ijt_n} + \theta^{\alpha}_{i,n+1}(t - t_{j1}) d_{ijt_1} + \cdots + \theta^{\alpha}_{i,2n}(t - t_{jn}) d_{ijt_n} + e^{\alpha}_{ijt}$<br>$\beta_{it,k} = \theta_{ik,0} d_{ijt_0} + \theta_{ik,1} d_{ijt_1} + \cdots + \theta_{ik,n} d_{ijt_n} + \theta_{ik,n+1}(t - t_{j1}) d_{ijt_1} + \cdots + \theta_{ik,2n}(t - t_{jn}) d_{ijt_n} + e_{it,k}$ | -12371 | 22046 |

[*] Computed using the importance sampling method of Newton and Raftery(1994). 5 percent of households randomly chosen from Sample 2 is used.

Table 3.4: Parameter Estimates (Posterior Standard Deviation)

| Satiation | Halls1 | Halls2 | Halls3 | Halls4 | Halls5 | Halls6 | Halls7 | Ricola1 | Ricola2 |
|---|---|---|---|---|---|---|---|---|---|
| $E\left[\alpha_{ijt}\right]$(std.dev) | 0.571 (0.005) | 0.552 (0.005) | 0.572 (0006) | 0.553 (0.008) | 0.570 (0.008) | 0.551 (0.010) | 0.564 (0.012) | 0.563 (0.007) | 0.551 (0.01) |
| Baseline | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ |
| $E\left[\beta_{it}\right]$(std.dev) | 0.494 (0.004) | 0.515 (0.008) | 0.507 (0.004) | 0.518 (0.010) | 0.512 (0.006) | 0.502 (0.004) | 0.526 (0.007) | 0.527 (0.009) | 0.540 (0.016) |

Note: $\frac{\alpha_{ijt}}{1-\alpha_{ijt}} = exp\left(\alpha^*_{ijt}\right)$, $\psi_{ijt} = exp(X_j^{'}\beta_{it})$

either C1 or C2 but not both, and either C2 or C3, but not both. Estimates of parameters in Model 5 are provided in Appendix A. Specifically, Tables A8 and A9 provide estimates of baseline preference for Halls and Ricola, and Tables A10 and A11 provide estimates of satiation parameter for Ricola. Across both brands, estimates of coefficients on past choices in both baseline preferences and satiation preferences are increasing after the first purchase of the brand. I find that coefficient for no experience, which could be interpreted as the prior, and coefficient for the first purchase relating to both baseline preferences and satiation parameters are statistically significant.

Figure 2 summarizes the distributions of estimates on past purchases of Halls. The box plot in Figure 2 shows the relationship between the distribution of the satiation parameter and the number of experience. The distribution of satiation parameter shifts upwards after the first purchase of Halls, and then it shifts downwards over experiences. Since a low level of satiation parameter represents a purchase of multiple items in the model, the changes of the distribution in Figure 2 shows that consumers tend to become variety-seekers after two purchases of Halls. The distribution of coefficients for past choices in Figure 2 shows that the density for no purchase is significantly different from densities for the other number of purchases. Mean or median of the coefficients related to any number of experience is higher than that of the coefficient for no experience.

To test whether the coefficient for no experience is less than the coefficient for the number of purchases, I conduct a Wilcoxon signed-ranks test, which is a non-parametric equivalent of the paired t-test[94]. I formally test whether brand preference changes before and after an additional purchase of a brand on the same consumer. It rejects the null hypothesis that both the distribution of coefficient for no purchase and the distribution of coefficient for one-time purchase are the same at any level. T-test assuming normal distribution also rejects the null hypothesis that mean of the coefficient for no experience and mean of the coefficient for one-time purchase is the same at any level. With a two-sided test and one-sided test, I conclude that the mean of the coefficient for one-time purchase is greater than the mean of the coefficient for no experience. Also, I conduct a sign test that tests the equality of matched pairs of the coefficients for no brand and coefficients for a one-time purchase. (See [**?** 88].) The null hypothesis is that the median of the differences between coefficients for no brand and coefficients for one-time purchase is zero; no further assumptions are made about the distributions. This result of the one-sided test rejects the null hypothesis significantly at any level, while the alternative hypothesis is that the median of differences between coefficients for no brand and coefficients for one-time purchases is greater than zero. With these tests, I compare coefficients for the one-time purchase, two-time purchase, three-time purchase, four-time purchase, with the five-time purchase and conclude that there are no significant differences after the one-time purchase.

This demonstrates that the degree of state dependence does not change after the first experience, and consumers learn most of the information at the first experience of a brand. When I compare the coefficients of no experience of a brand with the coefficients of the last purchase of a brand within a household's estimates, 73.2% of households shows that satiation parameters have increased after purchasing a brand. Furthermore, in comparison with the estimates of the one-time purchase of a brand.49% households have larger estimates of the last purchase of a brand.

## 3.6 Robustness Checks

### *3.6.1 Brand Choice Dependence or Reinforced Flavor Preference?*

In this cough drop market, consumers who prefer eucalyptus, menthol, or honey flavors are likely to choose Halls rather than Ricola. In contrast, consumers who prefer herb, cherry, and honey-lemon flavors are likely to choose Ricola. The structural dependence could be derived from the reinforcement of flavor preference (e.g., [39]). For $k$=*3,..,9*, $\theta_{ik,n}$ shows whether preference on each flavor is reinforced by experience or not. Most coefficients on past purchases in the baseline parameter are statistically insignificant except for one flavor. In summary, the preference for each flavor is not changed by the brand experience, and auto-correlated taste shock does not lead to state dependence.

Figure 3.1: Distribution of Satiation Parameter Estimates for Halls

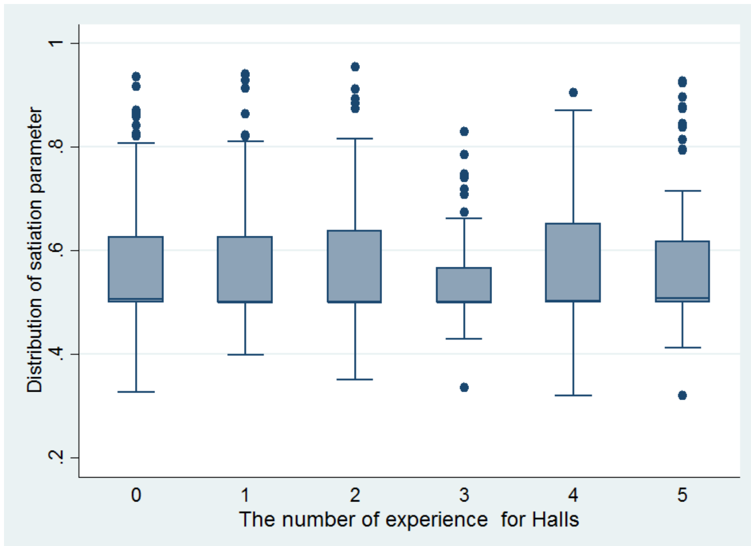
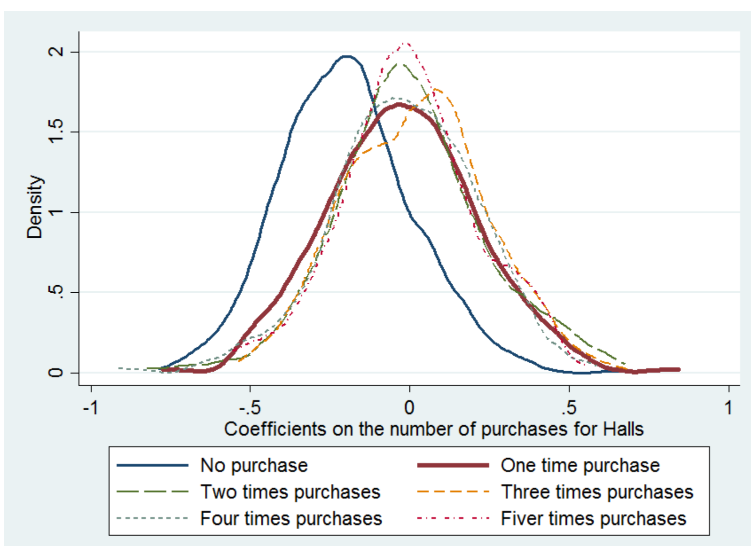



Figure 3.2: Distribution of a History of Past Choices Coefficients for Halls



### *3.6.2 Different Econometric Models*

As an exercise, I also ran the multinomial logit random effect model ignoring variety-seeking behavior.[8] The results are also consistent with the learning model, not the switching costs model. The model includes the following control variables: change of store, female head age, male head education level, male head job, quarterly purchase frequency, seasonality, and trend.

### *3.6.3 Change of Store, and Purchase Frequency*

By adding a dummy variable representing a change of store, I control for the possibility of brand switching resulting from a change of store. Since consumers are likely to form a new consideration-set when they change the store, consumers are likely to switch brands. Also, consumers with a high frequency of visits are likely to form a habit of choosing a specific brand; thus, they have a higher degree of state-dependence than consumers with a low frequency of visit. However, there is no significant statistical association between the frequency of visits and brand preference.

### *3.6.4 Searching*

It is likely that high searching costs could lead consumers to purchase products they have already purchased. Also, consumers who are working fewer hours or have flexible schedules are likely to try a search because of low searching costs. Therefore, I use the following proxy variables for search cost: the retiree, housewife, househusband, and low working hour jobs. The results still show state-dependence and provide evidence on learning, and I conclude that the main sources of state dependence are not searching costs.

[8]The results are available upon request

### 3.7 Conclusion

An important question in the fields of economics and marketing is what mechanism generates dynamics in consumer choice. Three competing theories of this mechanism, learning model, switching costs model, and satiation in variety-seeking behavior, are all suggested in previous studies. Each mechanism leads to a different history of brand choice. For example, the switching costs model predicts persistent choice, while satiation in a variety-seeking behavior model implies frequent brand switching or simultaneous demand for varieties. On the other hand, the learning model predicts that consumers conduct strategic trials of a product in the initial period and then show persistent choice thereafter. Hence, each model suggests different implications for managerial policy and optimal pricing policy for firms such as pricing, new product releases, expenditure on advertising, and research and development for current and future profit. This paper provides new methods for identifying the true source of dynamics and suggests evidence of learning behavior.

The application of these methods to cough drop markets addresses three questions:1) Does past choice affect the present choice? 2) If so, do all previous purchases affect current choice, as in Bayesian learning, or does the most recent purchase affect present choice, as in switching costs? 3)Finally, after controlling for satiation in variety-seeking behavior, does satiation lead to brand switching?

I find, first, strong evidence on structural state dependence. Both observed persistent brand choice and brand switching depend on past choices, even after controlling for consumers' unobserved heterogeneity and the decaying effect of each purchase. Second, this state dependence depends on only the first two purchases of a brand at most. This means that the uncertainty of a brand is resolved with only one or two experiences, and consumers learn fast. Also, consumers are likely to choose a larger size when they choose a brand they have already experienced and a smaller size when they switch to other brands not already experienced. This pattern of size choice also supports the decreasing cost of uncertainty to experience. Moreover, price elasticity or sensitivity grows higher as consumers become experienced. Finally, I find that only 11% of transactions are related to the purchase of multiple items, and the same product satiates 32% of consumers over time. This satiation behavior is derived from characteristics of a product such as a flavor, rather than brand loyalty. Therefore, even after controlling for a variety-seeking behavior, a consumer's present choice still depends on the first one or two purchases of the brand.

# BIBLIOGRAPHY

[1] Daniel A. Ackerberg. Advertising, learning, and consumer choice in experience good markets: An empirical examination. *International Economic Review*, 44(3):1007–1040, 2003.

[2] Sumit Agarwal, John C. Driscoll, Xavier Gabaix, and David Laibson. Learning in the credit card market. 2013.

[3] Susan Athey and Guido W. Imbens. Discrete choice models with multiple unobserved choice characteristics. *International Economic Review*, 48(4):1159–1192, 2007.

[4] Lynd Bacon and Peter Lenk. Augmenting discrete-choice data to identify common preference scales for inter-subject analyses. *Quantitative Marketing  Economics*, 10(4):453 – 474, 2012.

[5] Patrick Bajari, Phoebe Chan, Dirk Krueger, and Daniel Miller. A dynamic model of housing demand: Estimation and policy implications*. *International Economic Review*, 54(2):409–442.

[6] Badi H. Baltagi and Michael Keane. *Panel Data Discrete Choice Models of Consumer Demand*. Oxford University Press, 05 2015.

[7] Kapil Bawa. Modeling inertia and variety seeking tendencies in brand choice behavior. *Marketing Science*, 9(3):263–278, 1990.

[8] Alexandre Belloni, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017.

[9] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies*, 81(2):608–650, 11 2013.

[10] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50, May 2014.

[11] Alexandre Belloni, Victor Chernozhukov, and Ying Wei. Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34(4):606–619, 2016.

[12] David Besanko, Ulrich Doraszelski, and Yaroslav Kryukov. Is dynamic competition socially beneficial? the case of price as investment. *Retrieved from https://repository.upenn.edu/marketing*$_p$*apers*/325, 2016.

[13] Chandra R Bhat. The multiple discrete-continuous extreme value (mdcev) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, 42(3):274–303, 2008.

[14] Thomas Blake, Chris Nosko, and Steven Tadelis. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174, 2015.

[15] Christopher Bliss. A theory of retail pricing. *The Journal of Industrial Economics*, 36(4):375–391, 1988.

[16] Jack W Brehm. Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, 52(3):384, 1956.

[17] Bart J. Bronnenberg, Sanjay K. Dhar, and Jean-Pierre H. Dubé. Brand history, geography, and the persistence of brand shares. *Journal of Political Economy*, 117(1):87–115, 2009.

[18] Dubé Jean-Pierre Bronnenberg, Bart J. and Matthew Gentzkow. the evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6), 2012.

[19] Nuno Camacho, Bas Donkers, and Stefan Stremersch. Predictably non-bayesian: Quantifying salience effects in physician learning about drug quality. *Marketing Science*, 30(2):305–320, 2011.

[20] Gordon P. Wright Carrie M. Heilman, Douglas Bowman. The evolution of brand preferences and choice behaviors of consumers new to a market. *Journal of Marketing Research*, 37(2):139–155, 2000.

[21] Tat Chan, Chakravarthi Narasimhan, and Ying Xie. Treatment effectiveness and side effects: A model of physician learning. *Management Science*, 59(6):1309–1325, 2013.

[22] Hai Che, Tülin Erdem, and T. Sabri Öncü. Consumer learning and evolution of consumer brand preferences. *Quantitative Marketing and Economics*, 13(3):173–202, 2015.

[23] Hai Che, K Sudhir, and P.B Seetharaman. Bounded Rationality in Pricing Under State-Dependent Demand: Do Firms Look Ahead, and if So, How Far? *Journal of Marketing Research*, 44(3):434–449, 2007.

[24] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

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

[26] Andrew T. Ching, TÃlin Erdem, and Michael P. Keane. Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013.

[27] Andrew T. Ching, TÃlin Erdem, and Michael P. Keane. A simple method to estimate the roles of learning, inventories and category consideration in consumer choice. *Journal of Choice Modelling*, 13:60 – 72, 2014. Special Issue on Multi-stage Choice Models.

[28] Andrew T. Ching, Susumu Imai, Masakazu Ishihara, and Neelam Jain. A practitioner's guide to bayesian estimation of discrete choice dynamic programming models. *Quantitative Marketing and Economics*, 10(2):151–196, 2012.

[29] Andrea Coscelli and Matthew Shum. An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*, 122(2):213 – 246, 2004.

[30] Gregory S. Crawford and Matthew Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

[31] Jean-Pierre Dubé, Günter J Hitsch, and Peter E Rossi. Do switching costs make markets less competitive? *Journal of Marketing Research*, 46(4):435–445, 2009.

[32] Jean-Pierre Dubé, Günter J Hitsch, Peter E Rossi, and Maria Ana Vitorino. Category pricing with state-dependent utility. *Marketing Science*, 27(3):417–429, 2008.

[33] Jean-Pierre Dubé, J Hitsch, and Peter E Rossi. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445, 2010.

[34] Janine Empen and Stephen F. Hamilton. How do supermarkets respond to brand-level demand shocks? Evidence from the German beer market. *American Journal of Agricultural Economics*, 95(5):1223–1229, 2013.

[35] Tülin Erdem. A dynamic analysis of market structure based on panel data. *Marketing science*, 15(4):359–378, 1996.

[36] Tülin Erdem, Susumu Imai, and Michael P. Keane. Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, 1(1):5–64, 2003.

[37] Tülin Erdem, Michael L. Katz, and Baohong Sun. A simple test for distinguishing between internal reference price theories. *Quantitative Marketing and Economics*, 8(3):303–332, Sep 2010.

[38] Tülin Erdem, Joffre Swait, Susan Broniarczyk, Dipankar Chakravarti, Jean-Noel Kapferer, Michael Keane, John Roberts, Jan-Benedict EM Steenkamp, and Florian Zettelmeyer. Brand equity, consumer learning and choice. *Marketing Letters*, 10(3):301–318, 1999.

[39] TÃlin Erdem and Michael P. Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, 15(1):1–20, 1996.

[40] Tülin Erdem, Michael P. Keane, and Baohong Sun. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science*, 27(6):1111–1125, 2008.

[41] Ido Erev and Alvin E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88(4):848–881, 1998.

[42] Joseph Farrell and Paul Klemperer. Coordination and lock-in: Competition with switching costs and network effects. volume 3, chapter 31, pages 1967–2072. Elsevier, 1 edition, 2007.

[43] Nick Feltovich. Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. *Econometrica*, 68(3):605–641, 2000.

[44] Bram Foubert and Els Gijsbrechts. Try it, you'll like it—or will you? the perils of early free-trial promotions for high-tech service adoption. *Marketing Science*, 35(5):810–826, 2016.

[45] Brett R. Gordon and Baohong Sun. A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science*, 34(3):452–470, 2015.

[46] Gautam Gowrisankaran and Marc Rysman. Dynamics of Consumer Demand for New Durable Goods. *Journal of Political Economy*, 120(6):1173–1219, 2012.

[47] Peter M. Guadagni and John D. C. Little. A logit model of brand choice calibrated on scanner data: A 25th anniversary perspective. *Marketing Science*, 27(1):26–28, 2008.

[48] Ali Umut Guler, Kanishka Misra, and Naufel Vilcassim. Countercyclical pricing: A consumer heterogeneity explanation. *Economics Letters*, 122(2):343–347, 2014.

[49] Vassilis Hajivassiliou, Daniel McFadden, and Paul Ruud. Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics*, 72(1-2):85–134, 1996.

[50] Wesley R. Hartmann. Intertemporal effects of consumption and their implications for demand elasticity estimates. *Quantitative Marketing and Economics*, 4(4):325–349, Dec 2006.

[51] Wesley R. Hartmann. Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science*, 29(4):585–601, 2010.

[52] Shohei Hasegawa, Nobuhiko Terui, and Greg M. Allenby. Dynamic brand satiation. *Journal of Marketing Research*, 49(6):842–853, 2012.

[53] James J Heckman. Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press, 1981.

[54] Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006.

[55] Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006.

[56] Igal Hendel and Aviv Nevo. Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7):2722–2751, 2013.

[57] Günter J. Hitsch. An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science*, 25(1):25–50, 2006.

[58] Bing Jing. Social learning and dynamic pricing of durable goods. *Marketing Science*, 30(5):851–865, 2011.

[59] Michael P Keane. A computationally practical simulation estimator for panel data. *Econometrica: Journal of the Econometric Society*, pages 95–116, 1994.

[60] Michael P. Keane. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business Economic Statistics*, 15(3):310–327, 1997.

[61] Jaehwan Kim, Greg M. Allenby, and Peter E. Rossi. Modeling consumer demand for variety. *Marketing Science*, 21(3):229–250, 2002.

[62] Jaehwan Kim, Greg M. Allenby, and Peter E. Rossi. Product attributes and models of multiple discreteness. *Journal of Econometrics*, 138(1):208 – 230, 2007. 50th Anniversary Econometric Institute.

[63] Jung Seek Kim and Brian T. Ratchford. Consumer choice and use of multiple information sources for automobile purchases. *International Journal of Electronic Commerce*, 16(3):7 – 40, 2012.

[64] Paul Klemperer. Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The Review of Economic Studies*, 62(4):515–539, 1995.

[65] Song Lin, Juanjuan Zhang, and John R Hauser. Learning from experience, simply. *Marketing Science*, 34(1):1–19, 2014.

[66] Yuping Liu-Thompkins and Leona Tam. Not all repeat customers are the same: Designing effective cross-selling promotion on the basis of attitudinal loyalty and habit. *Journal of Marketing*, 77(5):21–36, 2013.

[67] Leigh McAlister. A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research*, 9(2):141–150, 1982.

[68] Leigh McAlister and Edgar Pessemier. Variety seeking behavior: An interdisciplinary review. *Journal of Consumer Research*, 9(3):311–322, 1982.

[69] Nitin Mehta, Surendra Rajiv, and Kannan Srinivasan. Role of forgetting in memory-based choice decisions: A structural model. *Quantitative Marketing and Economics*, 2(2):107–140, Jun 2004.

[70] Robert J. Meyer and Yong Shi. Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem. *Management Science*, 41(5):817–834, 1995.

[71] Eugenio J. Miravete and Ignacio Palacios-Huerta. Consumer inertia, choice dependence, and learning from experience in a repeated decision problem. *The Review of Economics and Statistics*, 96(3):524–537, 2014.

[72] Nickolay V Moshkin and Ron Shachar. The asymmetric information model of state dependence. *Marketing Science*, 21(4):435–454, 2002.

[73] Alvin Murphy. A dynamic model of housing supply. *American Economic Journal: Economic Policy*, 10(4):243–67, November 2018.

[74] Sridhar Narayanan and Puneet Manchanda. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, 28(3):424–441, 2009.

[75] Peter W Newberry. An empirical study of observational learning. *The RAND Journal of Economics*, 47(2):394–432, 2016.

[76] Matthew Osborne. Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics*, 9(1):25–70, Mar 2011.

[77] Peter E. Rossi, Greg M. Allenby, and Robert McCulloch. *Bayesian Statistics and Marketing*. New York: John Wiley  Sons, 2005.

[78] Rishin Roy, Pradeep K. Chintagunta, and Sudeep Haldar. A framework for investigating habits, âœthe hand of the past,â and heterogeneity in dynamic brand choice. *Marketing Science*, 15(3):280–299, 1996.

[79] Takuya Satomura, Jaehwan Kim, and Greg M Allenby. Multiple-constraint choice models with corner and interior solutions. *Marketing Science*, 30(3):481–490, 2011.

[80] Pasquale Schiraldi. Automobile replacement: A dynamic structural approach. *RAND Journal of Economics*, 42(2):266–291, 2011.

[81] P. B. Seetharaman. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science*, 23(2):263–271, 2004.

[82] P. B. Seetharaman, Andrew Ainslie, and Pradeep K. Chintagunta. Investigating household state dependence effects across categories. *Journal of Marketing Research*, 36(4):488–500, 1999.

[83] Oleksandr Shcherbakov. Measuring consumer switching costs in the television industry. *The RAND Journal of Economics*, 47(2):366–393, 2016.

[84] Jiwoong Shin, K Sudhir, Luís Cabral, Jean-Pierre Dube, Günter J Hitsch, and Peter E Rossi. Commentaries and rejoinder to" do switching costs make markets less competitive?". *Journal of Marketing Research*, 46(4):446–452, 2009.

[85] Sangwoo Shin, Sanjog Misra, and Dan Horsky. Disentangling preferences and learning in brand choice models. *Marketing Science*, 31(1):115–137, 2012.

[86] Matthew Shum. Does advertising overcome brand loyalty? evidence from the breakfast-cereals market. *Journal of Economics Management Strategy*, 13(2):241–272, 2004.

[87] Nikolai Vasil'evich Smirnov. On the probabilities of large deviations. *Mat. Sb*, 40:443–454, 1933.

[88] George W Snedecor and William G Cochran Cochran. Statistical methods, eight ed. *Ames: Iowa State Univ. Press Iowa*, 1989.

[89] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[90] K Sudhir and Nathan Yang. Exploiting the choice-consumption mismatch: A new approach to disentangle state dependence and heterogeneity. *Cowles Foundation Discussion Paper*, (1941), 2014.

[91] Sarang Sunder, V. Kumar, and Yi Zhao. Measuring the lifetime value of a customer in the consumer packaged goods industry. *Journal of Marketing Research*, 53(6):901–921, 2016.

[92] J Miguel Villas-Boas. Consumer learning, brand loyalty, and competition. *Marketing Science*, 23(1):134–145, 2004.

[93] Elizabeth J Warner and Robert B. Barsky. The timing and magnitude of retail store markdowns: Evidence from weekends and holidays. *Quarterly Journal of Economics*, 110(2):321–352, 1995.

[94] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.

## .1 Appendix A Determinants of Brand Switching

To understand the true source of dynamic brand choice, in chapter 1 and chapter 2, I test the switching costs model and learning model to explain the persistent brand choice. In this section, as exercises, I investigate the determinants of brand switching. By identifying what the triggers of brand switching are, it helps to understand whether satiation behavior also one of the state-dependent choices or not.

- Variety seeker, Satiation on brand
- H0-1. Trend of popular flavors
- H0-2. The higher price, more switch brand
- H0-3. The larger brand loyalty, less switch brand
- H0-4. Variety seeker is more likely to switch. Household size, the second brand vs. the third brand in the shopping history
- H0-5. When consumer buy similar level of size than before, they are less likely to switch
- H0-6. As searching cost is lower, in terms of income, working time, and job, more switch

### *.1.1 Data*

- 14,645 Households in the panel for 2006-2015, 84,487 obs.
- To treat consumer as a new consumer, when I observe consumer went shopping a year later at first time, drop the previous shopping data for that consumer.

- Brand choice: Halls, Ricola, Fisherman's friend, Robitussin, Smith Brothers, Best Health, and non-national brands.

### .1.2 Reduced-Form

Consumer i chooses product j at time t

$$BRSW_{it} = \omega + P_{ijt}\alpha + X_{ijt}\beta + Z_{it}\gamma + Time_{it} + \varepsilon_{ijt}, \tag{11}$$

Consumer i chooses product j at time t and the date of n th purchase of brand $j, d_n$

$$U_{ijt} = \omega_j + \alpha_h P_{ijt} + \gamma_h X_{ijt} + \gamma_h B_{ijt} + Time_{it} + \varepsilon_{ijt}, \tag{12}$$

$$B_{ijt} = ln\, a_{ij1}\delta_{jd_1}^{t-d_1} + ln\, a_{ij2}\delta_{j2}^{t-d_2} + \cdots + ln\, a_{ijn\delta_{ijn}}^{t-d_n} \tag{13}$$

$$= ln\, a_{ij1} + (t - d_1)\, ln\, \delta_{ij1} + \cdots + ln\ a_{ijn} + (t - d_n)\, ln\, \delta_{ijn}$$

- L.variable, variable, L.variable*variable

- P: flavor, total size, log relative price per count, deal, coupon value, throat/cough, sugar-free.

- X: total quantity of each brand purchased in the shopping history, choice of brand, the order of purchase in the consecutive purchase of each brand, the order of brand purchase in the shopping history.

- Z: variety seeker, quarterly shopping frequency, female head age, male head age, household income, household size, education, occupation, employment(working hours).

### .1.3 Method

- Training sample: 71,931 obs.,and Test sample: 12,556 obs.

- By using double machine selection method, 1033 variables to 183 variables

Table A1: Machine Learning Classification for Sample2 with Test Sample

| | Linear Prob. | Logit with RE | Probit with RE | Distributed Random Forest | Gradient Boost Machine | Deep Learning w.o. layers | Deep Learning with layers |
|---|---|---|---|---|---|---|---|
| Logloss | | | | 0.257 | 0.222 | 0.739 | 0.272 |
| AUC | | | | 0.970 | 0.964 | 0.943 | 0.958 |
| Error rate at 0 | 0.074 | 0.077 | 0.076 | 0.046 | 0.063 | 0.059 | 0.071 |
| Error rate at 1 | 0.413 | 0.388 | 0.398 | 0.132 | 0.128 | 0.146 | 0.138 |
| Total error rate | 0.175 | 0.170 | 0.173 | babypink 0.072 | 0.082 | 0.085 | 0.091 |

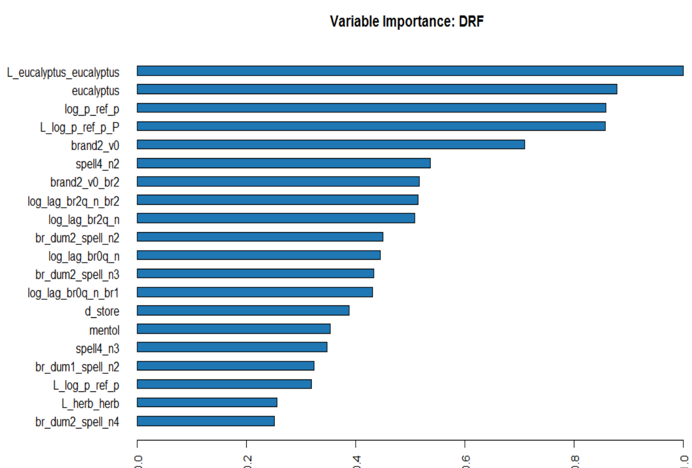


Table A2: Machine Learning Classification with Test Sample

| | Linear Prob. | Logit with RE | Probit with RE | Distributed Random Forest | Gradient Boost Machine | Deep Learning w.o. layers | Deep Learning with layers |
|---|---|---|---|---|---|---|---|
| Logloss | | | | 0.257 | 0.222 | 0.739 | 0.272 |
| AUC | | | | 0.970 | 0.964 | 0.943 | 0.958 |
| Error rate at 0 | 0.074 | 0.077 | 0.076 | 0.046 | 0.063 | 0.059 | 0.071 |
| Error rate at 1 | 0.413 | 0.388 | 0.398 | 0.132 | 0.128 | 0.146 | 0.138 |
| Total error rate | 0.175 | 0.170 | 0.173 | 0.072 | 0.082 | 0.085 | 0.091 |

### *.1.4 Variables*

- Training sample: 71,931 obs.,and Test sample: 12,556 obs.
- By using ML, 1033 variables to 183 variables

Table A3: The Result on the Regression of Brand Switching

| Dependent variable: Brand Switching | Linear Probability (1) | Random Effect Logit (2) | Random Effect Probit (3) |
|---|---|---|---|
| Satiation behavior | | | |
| The order of shopping trips | 0.003*** | 0.030*** | 0.016*** |
| | (0.000) | (0.002) | (0.001) |
| The first spell of brand choice | 0.027*** | 0.524*** | 0.278*** |
| | (0.004) | (0.042) | (0.023) |
| The second spell of brand choice | 0.196*** | 1.680*** | 0.935*** |
| | (0.005) | (0.043) | (0.024) |
| The third spell of brand choice | 0.196*** | 1.634*** | 0.899*** |
| | (0.006) | (0.049) | (0.027) |
| The fourth spell of brand choice | 0.224*** | 1.702*** | 0.944*** |
| | (0.007) | (0.059) | (0.033) |
| The second purchase | 0.158*** | 0.786*** | 0.464*** |
| in the consecutive purchase of nonnational brands | (0.007) | (0.052) | (0.030) |
| The third purchase | 0.199*** | 1.040*** | 0.602*** |
| in the consecutive purchase of nonnational brands | (0.010) | (0.076) | (0.044) |
| The fourth purchase | 0.256*** | 1.335*** | 0.774*** |
| in the consecutive purchase of nonnational brands | (0.016) | (0.109) | (0.063) |
| The second purchase | -0.265*** | -1.977*** | -1.044*** |
| in the consecutive purchase of Halls | (0.005) | (0.043) | (0.023) |
| The third purchase | -0.287*** | -2.841*** | -1.505*** |
| in the consecutive purchase of Halls | (0.005) | (0.071) | (0.035) |
| The fourth purchase | -0.263*** | -3.088*** | -1.606*** |
| in the consecutive purchase of Halls | (0.007) | (0.109) | (0.052) |
| The second purchase | -0.258*** | -2.060*** | -1.087*** |
| in the consecutive purchase of Ricola | (0.012) | (0.121) | (0.065) |
| The third purchase | -0.239*** | -2.313*** | -1.236*** |
| in the consecutive purchase of Ricola | (0.015) | (0.190) | (0.102) |
| The fourth purchase | -0.212*** | -2.769*** | -1.510*** |
| in the consecutive purchase of Ricola | (0.019) | (0.343) | (0.183) |
| The second purchase | -0.246*** | -1.585*** | -0.853*** |
| in the consecutive purchase of Fisherman's Friends | (0.031) | (0.277) | (0.152) |
| The second purchase | -0.362*** | -2.600*** | -1.474*** |
| in the consecutive purchase of Robitussin | (0.065) | (0.518) | (0.295) |
| The second purchase | -0.261*** | -1.523*** | -0.805*** |
| in the consecutive purchase of Smith Brothers | (0.035) | (0.280) | (0.152) |
| The second purchase | -0.238*** | -1.360*** | -0.745*** |
| in the consecutive purchase of Best Health | (0.041) | (0.299) | (0.162) |

··· Continued in the next page ···

Table A4: Continued:The Result on the Regression of Brand Switching

| Dependent variable: Brand Switching | Linear Probability (1) | Random Effect Logit (2) | Random Effect Probit (3) |
|---|---|---|---|
| Previous Brand Choice= Halls | 0.080*** | 0.909*** | 0.453*** |
| | (0.014) | (0.124) | (0.069) |
| Previous Brand Choice= Ricola | 0.062*** | 0.472** | 0.149 |
| | (0.023) | (0.228) | (0.124) |
| Previous Brand Choice=Fisherman's Friends | 0.285*** | 1.473*** | 0.774** |
| | (0.057) | (0.571) | (0.313) |
| Previous Brand Choice=Robitussin | 0.294** | 3.040** | 1.685** |
| | (0.127) | (1.332) | (0.718) |
| Previous Brand Choice=Smith Brothers | 0.173*** | 1.260** | 0.615** |
| | (0.062) | (0.551) | (0.304) |
| Previous Brand Choice=Best Health | 0.389*** | 2.136*** | 0.872** |
| | (0.080) | (0.749) | (0.376) |
| Learning behavior | | | |
| Log Total quantity accumulated for Halls | -0.113*** | -0.841*** | -0.466*** |
| *Previous Brand Choice= Halls | (0.002) | (0.022) | (0.012) |
| Log Total quantity accumulated for Ricola | -0.102*** | -0.710*** | -0.382*** |
| *Previous Brand Choice= Ricola | (0.004) | (0.048) | (0.026) |
| Log Total quantity accumulated for Fisherman's Friends | -0.134*** | -0.858*** | -0.476*** |
| *Previous Brand Choice= Fisherman's Friends | (0.011) | (0.117) | (0.064) |
| Log Total quantity accumulated for Robitussin | -0.121*** | -1.112*** | -0.621*** |
| *Previous Brand Choice= Robitussin | (0.033) | (0.322) | (0.173) |
| Log Total quantity accumulated for Smith Brothers | -0.094*** | -0.713*** | -0.391*** |
| *Previous Brand Choice= Smith Brothers | (0.014) | (0.120) | (0.067) |
| Log Total quantity accumulated for Best Health | -0.129*** | -0.845*** | -0.401*** |
| *Previous Brand Choice= Best Health | (0.019) | (0.173) | (0.087) |
| Log Total quantity accumulated for non-national brands | -0.098*** | -0.656*** | -0.369*** |
| *Previous Brand Choice= non-national brands | (0.002) | (0.021) | (0.012) |
| Log time elapsed from the previous purchase | 0.006*** | 0.057*** | 0.036*** |
| | (0.001) | (0.008) | (0.004) |

· · · Continued in the next page · · ·

Table A5: Continued:The Result on the Regression of Brand Switching

| Dependent variable: Brand Switching | Linear Probability (1) | Random Effect Logit (2) | Random Effect Probit (3) |
|---|---|---|---|
| Product Characteristics | | | |
| Eucalyptus flavor purchased at the previous shopping | 0.161*** | 0.967*** | 0.556*** |
| | (0.005) | (0.039) | (0.022) |
| Eucalyptus flavor purchased at the present shopping | 0.154*** | 0.959*** | 0.539*** |
| | (0.005) | (0.037) | (0.021) |
| Eucalyptus flavor purchased at the previous shopping | -0.330*** | -2.396*** | -1.337*** |
| *Eucalyptus flavor purchased at the present shopping | (0.006) | (0.053) | (0.030) |
| Mentol flavor purchased at the previous shopping | 0.121*** | 0.855*** | 0.494*** |
| | (0.007) | (0.053) | (0.030) |
| Mentol flavor purchased at the previous shopping | -0.225*** | -1.454*** | -0.835*** |
| *Mentol flavor purchased at the present shopping | (0.008) | (0.060) | (0.034) |
| Honey flavor purchased at the previous shopping | 0.007 | -0.038 | -0.030 |
| | (0.015) | (0.125) | (0.070) |
| Honey flavor purchased at the present shopping | 0.024 | 0.123 | 0.104 |
| | (0.015) | (0.126) | (0.070) |
| Honey flavor purchased at the previous shopping | -0.115*** | -0.838*** | -0.491*** |
| *Honey flavor purchased at the present shopping | (0.021) | (0.181) | (0.101) |
| Herb flavor purchased at the previous shopping | 0.118*** | 0.689*** | 0.393*** |
| | (0.011) | (0.088) | (0.049) |
| Herb flavor purchased at the present shopping | 0.214*** | 1.388*** | 0.780*** |
| | (0.009) | (0.080) | (0.044) |
| Herb flavor purchased at the previous shopping | -0.436*** | -3.723*** | -2.014*** |
| *Herb flavor purchased at the present shopping | (0.014) | (0.154) | (0.081) |
| Cherry flavor purchased at the previous shopping | 0.007 | 0.082** | 0.052** |
| | (0.005) | (0.039) | (0.022) |
| Cherry flavor purchased at the present shopping | -0.007 | -0.006 | 0.000 |
| | (0.005) | (0.039) | (0.022) |
| Cherry flavor purchased at the previous shopping | -0.006 | -0.218*** | -0.135*** |
| *Cherry flavor purchased at the present shopping | (0.007) | (0.062) | (0.035) |
| Lemon flavor purchased at the previous shopping | 0.017 | 0.224* | 0.142** |
| | (0.015) | (0.126) | (0.071) |
| Lemon flavor purchased at the present shopping | 0.003 | 0.087 | 0.025 |
| | (0.015) | (0.127) | (0.071) |
| Lemon flavor purchased at the previous shopping | 0.064*** | 0.349* | 0.194* |
| *Lemon flavor purchased at the present shopping | (0.021) | (0.183) | (0.103) |
| Fruit flavor purchased at the previous shopping | 0.009 | 0.103 | 0.054 |
| | (0.011) | (0.094) | (0.053) |
| Fruit flavorpurchased at the present shopping | 0.015 | 0.164* | 0.100** |
| | (0.010) | (0.093) | (0.050) |
| Fruit flavor purchased at the previous shopping | -0.078*** | -1.769*** | -0.807*** |
| *Fruit flavor purchased at the present shopping | (0.026) | (0.436) | (0.194) |

⋯ Continued in the next page ⋯

Table A6: Continued:The Result on the Regression of Brand Switching

| Dependent variable: Brand Switching | Linear Probability (1) | Random Effect Logit (2) | Random Effect Probit (3) |
|---|---|---|---|
| Product Characteristics | | | |
| I(L.total size$<$25) | -0.008 | 0.355* | 0.232** |
| | (0.024) | (0.209) | (0.117) |
| I(L.total size$<$25)*I(total size$<$25) | -0.096*** | -0.530** | -0.301** |
| | (0.024) | (0.209) | (0.117) |
| I(L.total size$<$25){*}I(24$<$total size$<$50) | -0.050** | -0.268 | -0.181 |
| | (0.023) | (0.198) | (0.112) |
| I(L.total size$<$25){*}I(49$<$total size$<$99) | -0.028 | -0.118 | -0.080 |
| | (0.024) | (0.210) | (0.118) |
| I(24$<$L.total size$<$50) | -0.021** | 0.308*** | 0.187*** |
| | (0.011) | (0.101) | (0.056) |
| I(24$<$L.total size$<$50)*I(total size$<$25) | 0.012 | 0.167* | 0.080 |
| | (0.011) | (0.095) | (0.054) |
| I(24$<$L.total size$<$50){*}I(24$<$total size$<$50) | -0.044*** | -0.213*** | -0.122*** |
| | (0.009) | (0.070) | (0.040) |
| I(24$<$L.total size$<$50){*}I(49$<$total size$<$99) | -0.002 | 0.047 | 0.021 |
| | (0.009) | (0.073) | (0.042) |
| I(49$<$L.total size$<$99) *I(total size$<$25) | 0.003 | 0.457*** | 0.253*** |
| | (0.010) | (0.098) | (0.054) |
| I(49$<$L.total size$<$99) {*}I(24$<$total size$<$50) | 0.031** | 0.343*** | 0.181*** |
| | (0.014) | (0.121) | (0.068) |
| I(49$<$L.total size$<$99){*}I(24$<$total size$<$50) | 0.001 | 0.077 | 0.053 |
| | (0.009) | (0.075) | (0.042) |
| I(49$<$L.total size$<$99){*}I(49$<$total size$<$99) | -0.026*** | -0.170** | -0.095** |
| | (0.009) | (0.074) | (0.042) |
| I(98$<$L.total size){*}I(total size$<$25) | 0.123*** | 1.212*** | 0.704*** |
| | (0.023) | (0.203) | (0.112) |
| I(98$<$L.total size){*}I(24$<$total size$<$50) | 0.068*** | 0.880*** | 0.506*** |
| | (0.011) | (0.099) | (0.055) |
| I(98$<$L.total size){*}I(49$<$total size$<$99) | 0.041*** | 0.751*** | 0.425*** |
| | (0.010) | (0.097) | (0.053) |

··· Continued in the next page ···

Table A7: Continued:The Result on the Regression of Brand Switching

| Dependent variable: Brand Switching | Linear Probability (1) | Random Effect Logit (2) | Random Effect Probit (3) |
|---|---|---|---|
| Product Characteristics | | | |
| Sugar Free product purchased at the previous shopping | -0.008 | -0.082* | -0.038 |
| | (0.005) | (0.043) | (0.024) |
| Sugar Free product purchased at the present shopping | -0.017*** | -0.124*** | -0.069*** |
| | (0.005) | (0.043) | (0.024) |
| Sugar Free product purchased at the previous shopping | 0.004 | -0.052 | -0.039 |
| *Sugar Free product purchased at the present shopping | (0.008) | (0.065) | (0.037) |
| Product for throat purchased at the previous shopping | 0.078*** | 0.486*** | 0.271*** |
| | (0.011) | (0.091) | (0.051) |
| Product for throat purchased at the present shopping | 0.184*** | 1.149*** | 0.649*** |
| | (0.010) | (0.080) | (0.045) |
| Product for throat purchased at the previous shopping | -0.390*** | -2.808*** | -1.542*** |
| *Product for throat purchased at the present shopping | (0.018) | (0.171) | (0.093) |
| Market Characteristics | | | |
| Change of Store | 0.041*** | 0.349*** | 0.196*** |
| | (0.003) | (0.025) | (0.014) |
| Demographic Characteristics | | | |
| Household Size | 0.004*** | 0.026* | 0.016** |
| | (0.001) | (0.014) | (0.008) |
| Female Head Age | 0.003** | 0.019* | 0.011* |
| | (0.001) | (0.012) | (0.007) |
| Variety Seeker: consumer who purchased | 0.032*** | 0.340*** | 0.212*** |
| at least two products at a time | (0.003) | (0.030) | (0.017) |
| I(Quarterly shopping frequency=2) | 0.024*** | 0.180*** | 0.115*** |
| | (0.004) | (0.031) | (0.017) |
| I(2$<$Quarterly shopping frequency$<$9) | 0.030*** | 0.215*** | 0.140*** |
| | (0.004) | (0.036) | (0.020) |
| I(8$<$Quarterly shopping frequency) | 0.013 | 0.117 | 0.089* |
| | (0.009) | (0.095) | (0.052) |
| Higher price paid than before | 0.252*** | 1.559*** | 0.877*** |
| | (0.008) | (0.068) | (0.037) |
| Constant | 0.634*** | 0.475 | 0.147 |
| | (0.136) | (1.217) | (0.708) |
| $\ln \sigma_\nu^2$ | | -0.936*** | -2.050*** |
| | | (0.058) | (0.056) |
| Observations | 71718 | 71718 | 71718 |
| AIC | . | 53332.271 | 53617.642 |
| BIC | . | 55104.106 | 55389.478 |

Note: The following variables are included: Household income, female head education, female head employment, female head occupation, dummy variable representing the relative price paid is higher than the previous relative price paid, a purchase with a deal, coupon value, log relative price.

Table A8: Internal Validity of Samples

| Samples Used in | Original Nielsen Data | Sample | Sample for Multiplediscreteness model |
|---|---|---|---|
| | | Mean (Median) | |
| Obs. | 356,686 | 127,581 | 38,163 |
| Total number of households | 40,383 | 28,724 | 11,139 |
| Household income | 19.82 (21) | 19.97 (21) | 20.09(21) |
| Household size | 2.44 (2) | 2.34 (2) | 2.36(2) |
| Household composition | 2.26 (1) | 2.31 (1) | 2.21(1) |
| Age and presence of children | 7.76 (9) | 7.95 (9) | 8.01(9) |
| Male head age | 5.56 (7) | 5.67 (7) | 5.87(8) |
| Female head age | 6.65 (8) | 6.89 (8) | 7.05(8) |
| Male head employment | 3.86 (3) | 3.89 (3) | 4.04(3) |
| Female head employment | 5.11 (3) | 5.14 (3) | 5.23(3) |
| Male head education | 3.08 (3) | 3.05 (3) | 3.08(3) |
| Female head education | 3.79 (4) | 3.79 (4) | 3.78(4) |
| Male head occupation | 6.41 (6) | 6.62 (6) | 6.87(6) |
| Female head occupation | 7.23 (8) | 7.79 (8) | 7.44(8) |
| Marital Status | 1.67 (1) | 1.68 (1) | 1.64(1) |
| Race | 1.30 (1) | 1.29 (1) | 1.31(1) |
| Time elapsed from the previous purchase | 173.18 (48) | 373.45 (203) | 462.34(337) |
| Total number of purchases per households | 8.83 (5) | 4.48 (2) | 8.07 (5) |
| For the households who went shopping at least 4 times, | | | |
| Obs. | 319,768 | 97,254 | 4,579 |
| No. of spells per households | 2.51 (2) | 3.22 (3) | 2.76(2) |
| Spell length for all spells | 2.58 (1) | 2.87 (2) | 2.20(2) |
| Fraction of brand switching per households | 0.40 (0.4) | 0.31 (0) | 0.106 (0) |
| FBS after first purchase of Halls | 0.30 (0) | 0.36 (0) | 0.13 (0) |
| FBS after non first purchase of Halls | 0.21 (0) | 0.22 (0) | 0.10 (0) |
| Obs. (Market share for years 2011-2015 in each sample) | | | |
| Halls | 102,043 (55.79%) | 51,686 (54.84%) | 34,480 (82.94%) |
| Ricola | 15,530 (8.49%) | 9,647 (10.24%) | 7,091 (17.06%) |
| Nonnational Brands | 65,329 (35.72%) | 32,911 (34.92%) | |

Table A9: The Effect of Lagged Brand Switching with Deal on Current Brand Switching

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| | Linear Probability | Logit | Probit |
| L.deal*L.brand switching | 0.055*** | 0.303*** | 0.168*** |
| | (0.007) | (0.049) | (0.028) |
| L.deal | -0.018** | -0.144** | -0.081** |
| | (0.008) | (0.060) | (0.034) |
| L.brand switching | 0.156*** | 0.853*** | 0.513*** |
| | (0.004) | (0.026) | (0.015) |
| Constant | 0.065*** | -3.360*** | -1.932*** |
| | (0.017) | (0.142) | (0.080) |
| Var($\mu_{0i}$) | | -0.488*** | -1.637*** |
| | | (0.046) | (0.046) |
| Observations | 79705 | 79705 | 79705 |
| AIC | . | 66666.593 | 66410.797 |
| BIC | . | 67892.356 | 67636.560 |

Controls are lagged variables, interaction terms with lagged variables and current variables for the following variables: log relative price, dummies of flavors, sugar-free, function on throat, log total quantity, change of store, female head age, male head education level, dummies on male head job, quarterly purchase frequency, variety, seasonality, trend. Var($\mu_{0i}$) is a random part and it shows heterogeneity of consumers preference

### *.1.5 Size transition process*

Since total size per bag varies across brands and stores, I transform size variable into a categorical variable as follows:

$$size_{i,t} = \begin{cases} 1 \quad if & \quad total\, counts\, consumed \quad < 25 \\ 2 \quad if \quad 24 < & total\, counts\, consumed \quad < 50 \\ 3 \quad if \quad 49 < & total\, counts\, consumed \quad < 99 \\ 4 \quad if \quad 98 < & total\, counts\, consumed \end{cases}$$

$$\triangle size_{i,t+1} = \begin{cases} 0 \quad if \quad \mathrm{size}_{i,t+1} < size_{i,t} \\ 1 \quad if \quad \mathrm{size}_{i,t+1} = size_{i,t} \\ 2 \quad if \quad \mathrm{size}_{i,t+1} > size_{i,t} \end{cases}$$

## .2 Appendix B Multiple-Discreteness Model

From equation (13), I derive the regression of optimal quantity as follows:

$$ln\left(q_{ijt}+1\right)=\frac{1}{1-\alpha_{ijt}}ln\left(\alpha_{ijt}\psi_{ijt}\right)+\frac{1}{\alpha_{ijt}-1}ln\left(p_{ijtj}\right)+\frac{1}{\alpha_{ijt}-1}\left(\varepsilon_{0}-\varepsilon_{j}\right)$$

Table B2: Evidence on Time Varying Parameters

| | (1) | (2) |
|---|---|---|
| | log total quantity+1 | log total quantity+1 |
| One-time purchase | 0.117*** | 0.187*** |
| for Halls | (0.023) | (0.024) |
| Two-time purchase | 0.152*** | 0.273*** |
| for Halls | (0.026) | (0.030) |
| Three-time purchase | 0.215*** | 0.383*** |
| for Halls | (0.029) | (0.034) |
| Four-time purchase | 0.221*** | 0.432*** |
| for Halls | (0.033) | (0.039) |
| Five-time purchase | 0.221*** | 0.498*** |
| for Halls | (0.040) | (0.047) |
| Six-time purchase | 0.160*** | 0.487*** |
| for Halls | (0.046) | (0.053) |
| Seven-time purchase | 0.175*** | 0.593*** |
| for Halls | (0.051) | (0.060) |
| Eight-time purchase | 0.136** | 0.583*** |
| for Halls | (0.058) | (0.068) |
| Nine-time purchase | 0.165*** | 0.593*** |
| for Halls | (0.063) | (0.073) |
| Ten-time purchase | 0.806*** | 1.261*** |
| for Halls | (0.020) | (0.047) |
| log(price per counts) | -0.226*** | -0.225*** |
| *no purchase | (0.005) | (0.005) |
| log(price per counts) | -0.195*** | -0.195*** |
| *one time purchase | (0.006) | (0.006) |
| log(price per counts) | -0.188*** | -0.189*** |
| *two-time purchase | (0.007) | (0.007) |
| log(price per counts) | -0.173*** | -0.173*** |
| *three-time purchase | (0.008) | (0.008) |
| log(price per counts) | -0.174*** | -0.176*** |
| *four-time purchase | (0.009) | (0.009) |
| log(price per counts) | -0.176*** | -0.178*** |
| *five-time purchase | (0.012) | (0.012) |
| log(price per counts) | -0.202*** | -0.205*** |
| *six-time purchase | (0.014) | (0.014) |
| log(price per counts) | -0.204*** | -0.205*** |
| *seven-time purchase | (0.016) | (0.016) |
| log(price per counts) | -0.216*** | -0.219*** |
| *eight-time purchase | (0.018) | (0.018) |
| log(price per counts) | -0.207*** | -0.213*** |
| *nine-time purchase | (0.020) | (0.020) |
| log (Time elapsed from | | -0.014*** |
| the first purchase) | | (0.002) |
| log (Time elapsed from | | -0.008*** |
| the second purchase) | | (0.003) |
| log (Time elapsed from | | -0.008** |
| the third purchase) | | (0.003) |
| log (Time elapsed from | | -0.007* |
| the fourth purchase) | | (0.004) |
| log (Time elapsed from | | -0.011** |
| the fifth purchase) | | (0.005) |
| log (Time elapsed from | | -0.008 |
| the sixth purchase) | | (0.005) |
| log (Time elapsed from | | -0.016*** |
| the seventh purchase) | | (0.006) |
| log (Time elapsed from | | -0.003 |
| the eighth purchase) | | (0.007) |
| log (Time elapsed from | | 0.007 |
| the ninth purchase) | | (0.008) |
| Constant | 2.972*** | 2.992*** |
| | (0.028) | (0.028) |
| Observations | 63436 | 63436 |
| Within R-squared | 0.313 | 0.315 |
| Overall R-squared | 0.403 | 0.407 |
| Between R-squared | 0.449 | 0.451 |

The following control variables are included: quarterly purchase frequency, household size, change of store, marital status, household internet connection.

Table B3: Full Latent Variable Model

| Parameter related to $\beta_{it,1}$ Halls | Posterior Mean | Posterior Standard Error |
|---|---|---|
| Never purchased Halls ($\theta_{i0}$) | 0.413*** | (.0120) |
| One-time purchase of Halls ($\theta_{i1}$) | 0.471*** | (0.014) |
| Two-time purchase of Halls ($\theta_{i2}$) | 0.031* | (0.017) |
| Three-time purchase of Halls ($\theta_{i3}$) | 0.028 | (0.020) |
| Four-time purchase of Halls ($\theta_{i4}$) | -0.013 | (0.023) |
| Five-time purchase of Halls ($\theta_{i5}$) | -0.021 | (0.022) |
| Time elapsed from the first purchase of Halls($\theta_{i6}$) | 0.078*** | (0.006) |
| Time elapsed from the second purchase of Halls($\theta_{i7}$) | -0.006 | (0.010) |
| Time elapsed from the third purchase of Halls($\theta_{i8}$) | 0.003 | (0.013) |
| Time elapsed from the fourth purchase of Halls($\theta_{i9}$) | 0.004 | (0.014) |
| Time elapsed from the fifth purchase of Halls($\theta_{i10}$) | -0.017 | (0.015) |

* The 90% credible interval does not contain zero (two-sided)

**The 95% credible interval does not contain zero (two-sided)

***The 99% credible interval does not contain zero (two-sided)

Table B4: Full Latent Variable Model

| Parameter related to $\beta_{it,2}$ Ricola | Posterior Mean | Posterior Standard Error |
|---|---|---|
| Never purchased Ricola ($\theta_{i0}$) | 0.417*** | (0.011) |
| One-purchase of Ricola ($\theta_{i1}$) | 0.045*** | (0.014) |
| Two-time purchase of Ricola ($\theta_{i2}$) | 0.027 | (0.017) |
| Three-time purchase of Ricola ($\theta_{i3}$) | 0.006 | (0.020) |
| Four-time purchase of Ricola ($\theta_{i4}$) | -0.015 | (0.022) |
| Five-time purchase of Ricola ($\theta_{i5}$) | 0.004 | (0.023) |
| Time elapsed from the first purchase of Ricola($\theta_{i6}$) | 0.080*** | (0.006) |
| Time elapsed from the second purchase of Ricola($\theta_{i7}$) | -0.003 | (0.011) |
| Time elapsed from the third purchase of Ricola($\theta_{i8}$) | 0.012 | (0.012) |
| Time elapsed from the fourth purchase of Ricola($\theta_{i9}$) | -0.009 | (0.013) |
| Time elapsed from the fifth purchase of Ricola($\theta_{i10}$) | -0.008 | (0.015) |

* The 90% credible interval does not contain zero (two-sided)

**The 95% credible interval does not contain zero (two-sided)

***The 99% credible interval does not contain zero (two-sided)

Table B5: Full Latent Variable Model

| Parameter related to $\alpha_{it} - 1$ Halls | Posterior Mean | Posterior Standard Error |
|---|---|---|
| Never purchased Halls ($\theta_{i0}$) | -0.197*** | (0.0003) |
| One-time purchase of Halls ($\theta_{i1}$) | -0.016*** | (0.0003) |
| Two-time purchase of Halls ($\theta_{i2}$) | -0.0022*** | (0.0003) |
| Three-time purchase of Halls ($\theta_{i3}$) | -0.0004 | (0.0003) |
| Four-time purchase of Halls ($\theta_{i4}$) | -0.00003 | (0.0003) |
| Five-time purchase of Halls ($\theta_{i5}$) | 0.0012*** | (0.0003) |
| Time elapsed from the first purchase of Halls($\theta_{i6}$) | -0.042*** | (0.0001) |
| Time elapsed from the second purchase of Halls($\theta_{i7}$) | 0.002*** | (0.0002) |
| Time elapsed from the third purchase of Halls($\theta_{i8}$) | -0.00001 | (0.0002) |
| Time elapsed from the fourth purchase of Halls($\theta_{i9}$) | 0.002*** | (0.0002) |
| Time elapsed from the fifth purchase of Halls($\theta_{i10}$) | 0.002*** | (0.0002) |

* The 90% credible interval does not contain zero (two-sided)

**The 95% credible interval does not contain zero (two-sided)

***The 99% credible interval does not contain zero (two-sided)

Table B6: Full Latent Variable Model

| Parameter related to $\alpha_{it} - 1$ Ricola | Posterior Mean | Posterior Standard Error |
|---|---|---|
| Never purchased Ricola ($\theta_{i0}$) | -0.195*** | (0.010) |
| One-time purchase of Ricola ($\theta_{i1}$) | -0.007 | (0.011) |
| Two-time purchase of Ricola ($\theta_{i2}$) | -0.005 | (0.013) |
| Three-time purchase of Ricola ($\theta_{i3}$) | -0.037 | (0.014) |
| Four-time purchase of Ricola ($\theta_{i4}$) | -0.010 | (0.018) |
| Five-time purchase of Ricola ($\theta_{i5}$) | 0.014 | (0.018) |
| Time elapsed from the first purchase of Ricola($\theta_{i6}$) | -0.045*** | (0.005) |
| Time elapsed from the second purchase of Ricola($\theta_{i7}$) | 0.005 | (0.007) |
| Time elapsed from the third purchase of Ricola($\theta_{i8}$) | -0.004 | (0.010) |
| Time elapsed from the fourth purchase of Ricola($\theta_{i9}$) | 0.019 | (0.010) |
| Time elapsed from the fifth purchase of Ricola($\theta_{i10}$) | -0.002 | (0.010) |

* The 90% credible interval does not contain zero (two-sided)

**The 95% credible interval does not contain zero (two-sided)

***The 99% credible interval does not contain zero (two-sided)

Table B1: Prior Distributions

| Prior | Setting |
|---|---|
| $vec\left(\theta_i\right) \sim N\left(\overline{\theta}, A\right)$ | $\overline{\theta} = 0,\ A = I_{nh} \times 10^{-2}$ |
| $V_\beta \sim IW\left(v_{\beta 0}, V_0\right)$ | $v_0 = (J+2), V_0 = I_{nk} \times (J+2)$ |
| $vec\left(\theta_i^\alpha\right) \sim N\left(\overline{\theta}^\alpha, A^\alpha\right)$ | $\overline{\theta} = 0,\ A = I_{n\alpha} \times 10^{-2}$ |
| $V_\alpha \sim IW\left(v_0, V_{\alpha 0}\right)$ | $v_0 = (J+2), V_{\alpha 0} = I_{Jb} \times (J+2)$ |

## .3 *Appendix C Hybrid MCMC Algorithm*

As an MCMC sampler, a random walk Metropolis algorithm is used. Households (i=1,...,N) provided choice and quantity information ($q_{ijt}$) for J alternatives at each occasion t=1,...,T$_i$.

Heterogeneity is introduced into the model by specifying a random effects distribution for demand parameters. A random effects distribution is specified for past histories.

*Prior Distribution*

I summarize the prior distribution in Table B1, where $\theta_i$ is the coefficient matrix of state-dependence as I formally define and V is the covariance matrix of $e_{it}$ in equation (4) (i.e., $V_\beta = Cov\left[e_{it}\right],\ e_{it} = \left[e_{it,1}, \ldots, e_{it,K}\right]'$,$V_\alpha = Cov\left[e_{it}^\alpha\right],\ e_{it}^\alpha = \left[e_{it,1}^\alpha, \ldots, e_{it,J}^\alpha\right]'$)

*Conditional Posterior Distributions*

I run 10,000 MCMC iterations in all models. In all models, we used the last 10,000 iterations to estimate posterior distributions of the model parameters. $\beta_{it} = \theta_i' h_{it} + e_{it},\ e_{it} \sim N\left(0, V_\beta\right)$, $\alpha_{it} = \theta_i^\alpha h_{it}^{\alpha\prime} + e_{it}^\alpha,\ e_{it}^\alpha \sim N\left(0, V_\alpha\right)$ where $\beta_{it}$ is an $1 \times nk$ matrix whose rows contain each of the household-level parameter vectors. $h_{it}$ is a vector of the history of past choices on the $nh$ covariates which describe differences between units and time.

$\beta_{it} | q_{it}, \alpha_{it}, \theta_i, V_\beta$

The posterior is as follows.

$$p\left(\beta_{it}|q_{it},\alpha_{it},\theta_i,V_\beta\right) \propto L_{it}\left(\beta_{it}\right) \times det|V_\beta|^{-1/2} exp\left\{-\frac{1}{2}\left(\beta_{it}-\theta_i' h_{it}\right)' V_\beta^{-1}\left(\beta_{it}-\theta_i' h_{it}\right)\right\}$$

where $\theta_i = (\theta_{i0};\theta_{i1};\ldots;\theta_{i,10})$,

$\left[n=0,\ h_{it}' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\right]$,

$\left[n=1,\ h_{it}' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & t-t_1 & 0 & 0 & 0 & 0 \end{pmatrix}\right]$,

$\left[n=2,\ h_{it}' = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & t-t_1 & t-t_2 & 0 & 0 & 0 \end{pmatrix}\right]$,

$\left[n\geq 5,\ h_{it}' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & t-t_1 & t-t_2 & t-t_3 & t-t_4 & t-t_5 \end{pmatrix}\right]$.

The likelihood $L_{it}$ is the likelihood function expressed in Equation (8) for consumer $i = 1,...,N$ at shopping trip $t = 1,...,T_i$

Setting r(=1,...,R) to MCMC iterations, we use Metropolis-Hasting with a random walk algorithm, where $i = 1,...,N$ and $t = 1,...,T_i$.

Start with $\beta_{it}^0$

The RW Metropolis must be scaled in order to function efficiently,

Draw $\beta_{it}^{(r)} = \beta_{it}^{(r-1)} + \varepsilon,\ \varepsilon \sim N\left(0, s^2 H^{-1}\right) = N\left(0, 0.75I\right)$ where automatic tuning scheme $s = 0.75 \approx 2.93/\sqrt{dim\left(\beta_{it}\right)}$ and $H = E\left[\frac{\partial^2 log\, l}{\partial\beta\partial\beta'}\right]$ with increments having covariance $s^2 V^{(r)}$, where $s$ is scaling constant and $V^{(r)}$ is the current draw of $V$.

Compute the acceptance probability which is $min\left[\frac{p\left(\beta_{it}^{(r)}|q_{it},\alpha_{it},\theta_i,V_\beta\right)}{p\left(\beta_{it}^{(r-1)}|q_{it},\alpha_{it},\theta_i,V_\beta\right)}, 1\right]$

$\theta_i|\left\{\beta_{it}\right\}, V_\beta$

$$vec\left(\theta_i\right)|\left\{\beta_{it}\right\}, V_\beta \sim N\left[vec\left(\hat{\theta}_i\right), V_\beta \bigotimes \left(h'h + A\right)^{-1}\right]$$

where $\hat{\theta}_i = \left(h'h + A\right)^{-1}\left(h'\beta_i + A\bar{\theta}\right),\ h = \left(h_1,...,h_{T_i}\right)',\ \beta_i = \left(\beta_{i1},\ldots,\beta_{iT_i}\right)'$

$V_\beta|\left\{\beta_{it}\right\}, \left\{\theta_i\right\}$

A standard one-for-one draw

$V|\left\{\beta_{it}\right\}, \left\{\theta_i\right\} \sim IW\left(v_0 + \sum_{i=1}^N T_i, V_0 + S\right)$, where $S = \sum_{i=1}^N \sum_{t=1}^{T_i}\left(\beta_{it} - \theta_i' h_{it}\right)\left(\beta_{it} - \theta_i' h_{it}\right)'$

$\alpha_{it}|q_{it},\beta_{it},\theta_i^\alpha,V_\alpha$

$$p\left(\alpha_{it}|q_{it},\beta_{it},\theta_i^\alpha,V_\alpha\right) \propto L_{it}\left(\alpha_{it}\right) \times det|V_\alpha|^{-1/2} exp\left\{-\frac{1}{2}\left(\alpha_{it}-\theta_i^{\alpha\prime} h_{it}^\alpha\right)' V_\alpha^{-1}\left(\alpha_{it}-\theta_i^{\alpha\prime} h_{it}^\alpha\right)\right\}$$

where $\theta_i^\alpha = \left(\theta_{i0}^\alpha;\theta_{i1}^\alpha;\ldots;\theta_{i,11}^\alpha\right),\ h_{it}^{\alpha\prime} = [1, h_{it}']$

Setting r(=1,...,R) to MCMC iterations, we use Metropolis-Hasting with a random walk algorithm, where $i = 1, ..., N$ and $t = 1, ..., T_i$.

Draw $\alpha^{(r)} = \alpha^{(r-1)} + \varepsilon_\alpha$, $\varepsilon_\alpha \sim N\left(0, s_\alpha^2 H^{-1}\right), s_\alpha \approx 2.93/\sqrt{dim\left(\alpha_{it}\right)}$

Compute the acceptance probability which is $min\left[\frac{p\left(\alpha_{it}^{(r)}|q_{it},\{\beta_{it}\},\theta_i^\alpha,V_\alpha\right)}{p\left(\alpha_{it}^{(r-1)}|q_{it},\{\beta_{it}\},\theta_i^\alpha,V_\alpha\right)}, 1\right]$

$\theta_i^\alpha | \left\{\alpha_{it}\right\}, V_\alpha$

$vec\left(\theta_i^\alpha\right) | \left\{\alpha_{it}\right\}, V_\alpha \sim N\left[vec\left(\hat{\theta}_i^\alpha\right), V_\alpha \bigotimes \left(h^{\alpha\prime} h^\alpha + A^\alpha\right)^{-1}\right]$

where $\hat{\theta}_i^\alpha = \left(h^{\alpha\prime} h^\alpha + A^\alpha\right)^{-1}\left(h^{\alpha\prime}\alpha_i + A^\alpha\bar{\theta}^\alpha\right)$, $h^\alpha = \left(1, h_1, ..., h_{T_i}\right)'$, $\alpha_i = \left(\alpha_{i1}, \ldots, \alpha_{iT_i}\right)'$

$V_\alpha | \left\{\alpha_{it}\right\}, \left\{\theta_i^\alpha\right\}$

A standard one-for-one draw $V_\alpha | \left\{\alpha_{it}\right\}, \left\{\theta_i^\alpha\right\} \sim IW\left(v_0 + \sum_{i=1}^N T_i, V_{\alpha 0} + S_\alpha\right)$, where $S = \sum_{i=1}^N \sum_{t=1}^{T_i}\left(\alpha_{it} - \theta_i^{\alpha\prime} h_{it}^\alpha\right)\left(\alpha_{it} - \theta_i^{\alpha\prime} h_{it}^\alpha\right)'$