THE UNIVERSITY OF CHICAGO

DISAGGREGATE NETWORK EFFECTS ON TWO-SIDED PLATFORMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
XILIANG LIN

CHICAGO, ILLINOIS
DECEMBER 2017

Copyright © 2017 by Xiliang Lin

All rights reserved

To my parents and wife

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank Pradeep Chintagunta, Jean-Pierre Dubé, Ali Hortaçsu, and especially my principal advisor Günter Hitsch for very helpful guidance, critique, support, and patience. I thank Øystein Daljord, Bora Keskin, Sanjog Misra, Anita Rao, Robert Sanders, and Brad Shapiro for very helpful discussions and comments. I thank the Kilts-Nielsen Data Center at The University of Chicago Booth School of Business for providing the data for this dissertation project.

ABSTRACT

This dissertation empirically explores the mechanisms through which software variety affects hardware purchases in markets with a hardware-software structure. Software variety may provide value to consumers because (i) it allows them to find a product better matched to their preferences (heterogeneity), and (ii) it may satisfy their need for different products across consumption occasions (within-person demand for variety). Using household-level coffee purchases and Keurig machine adoption data, I first build and estimates a consumer demand model for coffee allowing both within-consumer demand for variety and cross-consumer heterogeneity. Repeated purchases and trip-level variations in product prices and availabilities identify the household preferences, which allows for more reliable estimates. With the preference estimates, I then calculate the value of consuming Keurig's coffee pods (K-Cups) for each household and week, which I call the option value of K-Cups in the paper. Finally, I directly link the option values to their Keurig machine adoption decisions via a dynamic discrete choice model. Estimation results indicate variety in K-Cups, particularly brand variety, increases the option values for households, and the higher option values, in turn, increases machine adoption. To understand the mechanisms through which K-Cup variety influence Keurig machine adoption, I simulate the counterfactual adoption rates if households (a) can only choose their favorite K-Cup brands or ground coffee, and (b) have a library of K-Cups to choose from but not their favorite brands. Simulation results show that adoption rate would be about 2/3 of the current level in the later case compared to 1/4 in the former scenario, and thus indicate demand for variety is the primary mechanism under which K-Cup variety is valuable to consumers. Moreover, I show without third-party brands, (i) Keurig adoptions would have been much lower at the end of 2013; and (ii) the median consumer welfare loss

conditional on the adoption of Keurig machines is large compared to their spending on coffee with substantial heterogeneity. Furthermore, Keurig-owned K-Cups revenue would have been more than 40% without third-party brands because of lower adoption base.

CHAPTER 1

INTRODUCTION

Two-sided platforms are common, for example, video game platforms such as Sony PlayStation 4 and Microsoft Xbox One, e-reading platforms such Amazon Kindle, and single-cup serving coffee platforms such as the Keurig platform. These platforms often exhibit a hardware (i.e. game consoles, e-readers, and coffee machines) and software (i.e. video games, e-books, and coffee) structure (Katz and Shapiro, 1985). The extant literature (e.g., Gandal et al. (2000); Dubé et al. (2010); and Lee (2013)) has shown the importance of software variety to hardware adoptions in certain markets. Indeed, many platforms allow third parties to produce compatible software because of the importance of variety to hardware adoption. However, some platforms, such as Altria's MarkTen's electronic cigarette platform, Nestlé's Nespresso coffee platform and Nestlé's Dolce Gusto coffee platform, are closed platforms, where only the hardware manufacturer produces the software and thus limiting variety. Why do some platforms allow third-party software provisions while others do not? The answer to this question, in part, depends on the mechanisms through which software variety functions to generate hardware adoption. The extant literature has not empirically explored these mechanisms. This dissertation project seeks to bridge this gap and provide new micro-level evidence for network effects using individual level Keurig machine adoption and coffee purchase data.

The single-cup-serving coffee market in the U.S. with household-level purchase data provides a unique opportunity to generate insights about two-sided platforms. The single-cup-serving coffee market has the essential features of a two-sided market, such as tying, licensing, and network effects. The single-cup-serving coffee machines and corresponding portion packs are tied and do not work across platforms. For example, Keurig's coffee pods (K-Cups) do not work in Tassimo machines, and vice versa. Therefore, to consume coffee, a consumer needs to purchase coffee pods compatible with her machine. Whereas some platforms only allow in-house coffee pods, the Keurig platform allows other roasters to produce K-Cups by paying a royalty fee and meeting specific requirements (a form of licensing). A third-party roaster's participation may increase adoptions of the Keurig machine and thus increase the customer base for the platform (the cross network effect). The increase in customer base may indirectly benefit other roasters already on the Keurig platform (the indirect network effect) due to a larger customer pool. Roasters also compete in K-Cups sales (direct competition effect). As such, studying the Keurig platform helps us understand not only the single-cup-serving coffee market (a multi-billion-dollar market) but also how the complex dynamics of two-sided markets interact and function in general.

On two-sided platforms, software variety is of value to consumers if consumers are either heterogeneous or variety-seeking. If consumers are heterogeneous, more variety may allow them to find a better match in the software market and thus higher adoption values. Similarly, if a consumer is variety seeking, more variety directly increases the consumer's utility and thus adoption value. In aggregate data, both consumer heterogeneity and variety seeking lead to a similar empirical observation - the high correlation between software provisions and hardware adoptions. As such, heterogeneity versus variety seeking is hard if not impossible to distinguish in aggregate data. This feature poses a significant challenge to empirical researchers interested in the mechanisms of network effects because most data on two-sided platforms are indeed at the aggregated market level. Moreover, measuring adoption value from software variety is often difficult, because the software on many platforms is durable (e.g., video games and ebooks). If the consumer purchases each software provision only once, we cannot rely on variations in relative prices and software availability across shopping trips to identify consumer preferences and thus the values of software variety without resorting to strong assumptions of strategic consumer behavior.

For this study, I put together a panel of household coffee purchases and machine adoptions based on the Nielsen Homescan and the Nielsen RMS (Retail Measurement Services) retail scanner data. The Homescan data have detailed information on coffee purchases for an annual panel of around 60,000 households from 2007 to 2013. The data also include Keurig machine purchases for about 4,000 households. Because K-Cups are only for use in Keurig machines, I can further infer household adoptions from K-Cups purchases. The Nielsen RMS database includes data on weekly store-level sales units and average prices at UPC (universal product code) level, and thus allow me to infer product availability and prices in different shopping trips.[1]

The assembled household level data has several advantages over aggregate data when conducting analyses. First, the household-level data allow me to incorporate both across-household preference heterogeneity and within-household variety seeking in the coffee demand model. Thus, I can empirically distinguish them in the estimation. Second, coffee is a repeatedly purchased good, which allows for preference estimation through variations in prices and availability of competing products between shopping trips. With estimates of preferences at the household level, I can better infer the value from software (coffee) variety compared with previous work. The data also offer a third unique advantage – the ability to link the machine adoption and coffee purchases directly at the household level. As such, I can link the estimates of household preferences for coffee to the adoption decision through the value from software variety, which helps me identify the effect of software variety on hardware adoption.

1. The RMS data covers around 50% of shopping trips by Homescan households. The product availability and prices are further supplemented by imputation on large retailers. Details are in section 3.2.

Although both consumer heterogeneity and variety seeking result in more adoptions when software variety increases, they lead to different business implications for the platform owning firm. Many hardware manufacturers also produce the software (e.g. Sony's PS4 gaming platform and Green Mountain Coffee Roaster's (GMCR's) Keurig platform). In many tied goods markets, the firm prices the hardware at cost or loss, and recoups the loss and harvests profits from the software sales. Therefore, the firm needs to carefully estimate the value of different consumers from the software market to price and maximize profits. As the firm often holds patents or property rights over the platform, it can decide whether to allow third-parties to produce compatible software. When consumers are heterogeneous, they consume the most preferred software. Consequently, the consumers attracted by the third parties are of lower value since selling the firm's software to them is difficult. By comparison, variety seeking consumers are of higher value since the firm can sell to them. Of course, the firm can charge third party royalty fees or licensing fees, and the exact of values of third parties would depend on the charged fees and the competition in the aftermarket. It nevertheless highlights how heterogeneity versus variety seeking may influence the value of customers to the firm, which in turn could affect its business decisions.

To understand consumer valuation for variety on the Keurig platform, I first summarize household coffee purchases before and after their Keurig machine adoption. The median Herfindahl-Hirschman Index (HHI, Herfindahl (1950) and Hirschman (1945)) of brand expenditure before the adoption is 4,913, which compares to 2,880 after the adoption. The large decline in the concentration of brands purchased indicate households purchase more variety post adoption, and their ground coffee purchases were very concentrated before adoption. Further analysis reveals this change is mainly due to increase in purchases of multiple brands/flavors in the same shopping occasion and the more frequent temporal brand switching across shopping trips. Many reasons other than variety seeking, such as more frequent promotions or deeper price discounts, could explain their purchase patterns. Therefore, I resort to a formal model of coffee demand to isolate those effects.

To accommodate the purchases of multiple brands in a given choice occasion, I use a modified version of multiple discreteness model (Kim et al., 2002; Bhat, 2008). In the model, I control for both unobservable household heterogeneity and observable heterogeneity such as household level behaviors and demographics. I estimate the coffee demand model using the Bayesian hierarchical framework Rossi et al. (2005) on a cluster. With the household-level preference estimates, I estimate the expected option value of Keurig's coffee pods (K-Cups) for each household and time period. I then directly link the option values to their Keurig machine adoption decisions via a dynamic discrete choice model. The results show households display higher levels of satiation, lower levels of state dependence, and more homogeneous preferences for K-Cup brands compared to ground coffee brands even after controlling for prices and product availabilities. The higher satiation parameter means the marginal utility for K-Cups declines faster than ground coffee than thus contributes to the purchase of multiple brands on the same shopping occasion. The lower inertia makes purchasing different brands easier. The more homogeneous preferences for K-Cups brands, likely due to similar appearance and co-advertising, also contribute to the choice of multiple brands and temporal switching behavior. I also find that households have a much higher valuation for K-Cups (on average, about \$3.3 for the first serving at weekly level) compared to ground coffee (about \$1.7). The differential valuation may explain household purchase K-Cups in 84.6% of their coffee purchase trips after adoption, even if the average per-serving price of K-Cups is much higher (\$0.594 v.s. \$0.158). The estimation results for the machine-adoption problem show adoption probability increases with the consumption value of K-Cup. Hence, brand variety increases Keurig machine adoption through the consumption value of K-Cups.

In the counterfactual analysis, I show the adoption rate of Keurig machines would have been about 50% lower if GMCR did not allow third-party brands on the Keurig platform, and both heterogeneity and variety seeking contributes to the values of variety and thus hardware adoptions. To gauge the indirect network effect of third-party roasters, I estimate the revenue changes to GMCR-owned brands when the third-party and licensed brands were not available, and find GMCR revenue would have been much lower if the third-party and licensed brands were not allowed. The decrease is mainly due to the reduction in the customer base. The household preference distribution moderates these effects. Conditional on adoptions and the pricing path of Keurig machines and K-Cups, the median welfare loss for consumers is 33% of their spending on coffee with substantial heterogeneity if third-party or licensed brands were not available on the Keurig platform. I plan to run more counterfactuals, particularly for brands values to the platform and pricing in contrast with aggregate demand models, to demonstrate the importance of separately identifying heterogeneity and variety-seeking.

This dissertation proceeds as follows: chapter 2 reviews relevant literature, chapter 3 describes the single-serving coffee market and the data used in this dissertation, and presents some model free evidence for within-household demand for variety in coffee, chapter 4 describes the model to be estimated, chapter 5 describes the estimation procedure and algorithm, section 6 presents the estimation results, chapter 7 presents the counterfactuals, and chapter 8 summarizes and discusses implications of the results.

CHAPTER 2

LITERATURE REVIEW

Several theoretical papers study the structure of two-sided platforms and more generally tied-goods markets including the early work by Oi (1971), Katz and Shapiro (1985), Katz and Shapiro (1986), Farrell and Saloner (1985) and Farrell and Saloner (1986), and later work by Caillaud and Jullien (2003), Rochet and Tirole (2003), Rochet and Tirole (2006), Armstrong (2006) and Chen et al. (2009). These papers often treat either side of the network as homogenous and participation utility depends on the number of participants on the other side. Many platforms exhibit this feature such as the communication networks. However, to many platforms, both the participant-specific match (heterogeneity) and the number of participants (variety seeking) matter for participation utility. This dissertation project builds on this promise and studies the roles of heterogeneity and variety seeking in generating network effects.

My work is closely related to empirical work on two-sided platforms and more broadly tied goods. Most empirical papers use aggregate data to study network effects (e.g., Gandal et al. (2000), Nair et al. (2004), Clements and Ohashi (2005), Dubé et al. (2010) and Lee (2013)), which limit their ability to explore the mechanisms of cross-network effects and establish direct link of the software variety to the hardware adoption. Hartmann and Nair (2010) use individual level data to study razors adoption and blades consumption, but their focus is not on variety seeking, since it is not of primary importance in the market they study. Li (2016) studies the intertemporal pricing problem of Amazon Kindle and ebooks with a combination of individual level data and aggregate data. As ebooks are durable, she models the demand at genre level, which limits her ability to explore the mechanisms of the cross-network effects.

The aggregation may also be inappropriate because the consumer heterogeneity is over consumer genre preferences.

Two recent papers that also study the single-serving coffee market are relevant to my research. Both of these studies however, use aggregate, and not household-level, data and are therefore, subject to the concerns noted previously. Chintagunta et al. (2016) looks at the market in Portugal where, unlike the US market, all the platforms are closed ones. The focus of that paper is on the potential licensing decision of the hardware manufacturers and not on distinguishing between the variety seeking behavior and consumer preference heterogeneity. Kong et al. (2016), on the other hand, focus on quantifying the benefits of tied goods to platform growth. Both papers assume aggregate demand specification (Berry et al., 1995) to account for consumer heterogeneity. Since heterogeneity is identified due to non-IIA substitution patterns from the multinomial logit model, and multiple discreteness of purchases for K-Cups may make this assumption inappropriate in the Keurig setting. Moreover, the assumed model doesn't allow variety seeking other than the latent utility draws. Thus, the random utility shocks generate the indirect network effects, which may not be appropriate as shown later in this dissertation.

This dissertation is also related to works on the role of variety to consumer demand. Hoch et al. (1999) argue consumers value variety because it offers better opportunity to find a match, and in the case of a missing match, the next best alternative. Broniarczyk et al. (1998) find eliminating "low-preference" items while holding category shelf space constant has virtually no impact on consumer assortment perceptions and store choices. Moreover, Gourville and Soman (2005) argue that the variety may backfire when too many varieties are offered, and the variety offered differs in many dimensions. Complexity increases consumers' mental load, and thus trigger them to choose other brands. Consumers may also display both inertia and variety seeking behavior, and Bawa (1990) and Chintagunta (1998) proposed models to accommodate both. Similarly, the coffee-purchase model in this dissertation allows both inertia and variety seeking behavior.

CHAPTER 3

INDUSTRY AND DATA DESCRIPTION

3.1 Single-Cup-Serving Coffee Industry

A single-cup-serving system often consists of a coffee machine and portion-packed coffee. With the system, a consumer can brew a single cup of coffee with the click of a button. Brewing normally finishes within a minute. After brewing, the consumer can just discard the portion pack. The system offers a convenient way to brew coffee and requires minimal cleaning afterward. In general, the portion packs are incompatible across single-cup-serving coffee machines (i.e., Tassimo discs do not fit into the Keurig Machine and vice versa).

Single-cup-serving coffee machines quickly rose to prominence in the home coffee consumption market in the last ten years. Based on NCA (2016), 28% of US households have single-cup-serving coffee machines in 2016 compared to 1% in 2005. In 2016, roughly 36.2% of the US household coffee expenditure was on single-cup-serving coffee compared to 45.8% on ground and whole bean coffee (hereafter, ground) (IRI, 2016). The single-cup-serving coffee brewing systems have revolutionized how households brew coffee at home (home segment[1]). Keurig is largely a monopoly in this market with 85.6% market share.[2]

Keurig adopters have the option to purchase third-party brands. This feature is distinctive from Nespresso (the most popular single-cup-serving coffee machine in Europe) as well as several other single-cup-serving coffee machines such as Senseo. Third-party roasters generally pay a royalty fee to GMCR, the parent company of

1. Home segment is defined relative to away from home segment such as office coffee.

2. Market share is based on Nielsen Homescan data of 61,097 households with projection factor weight; the unweighted market share is 90%.

Keurig,[3] with some notable exceptions. For example, in the conversion with an industry insider from Starbucks, he reveals that GMCR pays Starbucks a fee for Starbucks K-Cups distributed by GMCR.[4] By comparison, Nespresso and Senseo machines do not allow other roasters/brands.

3.2 Data Description and Household Identification

This paper uses the Nielsen Homescan data and the Nielsen RMS (Retail Measurement Services) retail scanner data. Both data sets are available for academic research purposes through a partnership between the Nielsen Company and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business.[5] I supplement these data with imputation to better capture households' choice behaviors and machine adoption status.

The Nielsen Homescan database has comprehensive data on household coffee purchases and, to a certain extent, coffee-machine purchases. Participating households scan their purchased items and enter the corresponding prices after each shopping trip, hence the name "Homescan". If a participating household shops at a store, which shares information with Nielsen, the store directly transmits the purchased items and prices to Nielsen. Through the combination of these two methods, Nielsen tracks household purchases and corresponding prices. The Homescan database has an annual panel of about 60,000 households across the United States from the year 2007 onward. Before 2007, the panel size was round 40,000.

3. After Keurig K-Cup patent expiration in 2012, third-party roasters gradually started to offer compatible portion packs without paying GMCR a royalty fee.

4. The specific amount of the fee is confidential.

5. Information on availability and access to the data is available at http://research.chicagobooth.edu/nielsen/.

Because Homescan households only record the purchases they made, competing product information such as prices and availability is lacking. Therefore, I supplement the Homescan data with Nielsen RMS data to provide competing product information (shopping environment). The Nielsen RMS database includes data on weekly store-level sales units and average prices at UPC (universal product code) level. The data covers 39000+ stores across the United States from 2006. Although the coverage is expansive, the RMS stores is neither a census nor a randomly selected sample of retail stores. The RMS data cover about 55% of the coffee expenditure by Homescan households. To future improve the coverage of shopping environment, I impute additional price and availability data on coffee products based on Homesan purchases. Because pricing is rather homogeneous across stores of the same retailer (Hitsch et al., 2017), the large number of panelists in Homescan allows for coffee price imputation at retailer level.[6] Overall, the shopping environment data cover about 80% of the overall coffee expenditure by Homescan households.

To summarize households' purchases before and after their Keurig machine adoptions, I first need to identity the adopters and their corresponding adoption dates. I treat households with at least two purchases of K-Cups as Keurig adopters.[7] This identification method leaves 14,656 adoption households in Nielsen Homescan between 2007 to 2013. Out of these households, I observe the machine purchases for 4,085 households. Based on data from households recording their machine purchases, households often make their first K-Cup purchase within a few days of their machine purchase. Hence, I use the first purchase date of K-Cups as the adoption date for those not observed making a Keurig machine purchase.[8]

6. I only do price imputation for large coffee retailers for reliabiilty. Appendix E describes the data imputation procedures.

7. Appendix D documents household identification details.

8. Because of the imputation, adoption doesn't mean a machine purchase for all households. However, as discussed, the machine purchase dates tend to be very close to the adoption date as

Information on competing products is relevant in model estimation, because I rely on it to infer households' preferences from their choice behaviors. On the other hand, I only need household purchases, not information on competing products, for basic summaries. As such, the model free analyses in Section 3.3, 3.4 and 3.5 included all coffee consuming households. This choice ensures the comprehensiveness of model free analyses whenever possible.

Because I need information on competing products for model estimation, I focus on the data from 2008 to 2013 in the models. Before 2008, few people adopted the Keurig machine, and the distribution channel wasn't well set up yet (GMCR, 2009). I further restrict my study to top 30 designated market areas (DMAs)[9] with the most number of coffee consumers. I drop the small DMAs because the scant data make inferring information on competing products difficult. For example, the Glendive, MT DMA had only 3 Homescan households in 2013. The 30 DMAs covers 39,401 coffee-comsuming households, or roughly half of the entire Homescan database.[10] I further drop some households for a lack of competing product information in their choice data. The final dataset include 34,249 households with 7,805 Keurig machine adopters.

Estimating consumer preferences relies on the repeated purchases of coffee, and a long panel of purchases is essential. In the data, the selected households made 680,261 coffee purchases, inclusive of K-Cup purchases, within the coverage of shopping environment data. Thus, on average, the model utilizes about 20 purchases to infer

imputed. Hence, given the relatively long panel of purchases, the imputation should have little impact on the results. If the time interval between the machine purchase and K-Cup purchase is very long, households are not considered adopters on their machine purchase date. See Appendix F for details. Appendix F documents the details of machine adoption date imputation.

9. DMAs are major media markets defined by Nielsen with the same media coverage. Between 2008-2013, the Nielsen Homescan data covers about 200 DMAs.

10. In the whole database, I identified 76,047 households as coffee consumers and 14,656 adopted the Keurig machine.

household-specific preferences. The 7,805 adoption households made 110,608 purchases after their Keurig machine adoptions (on average 14.17 purchases/household), for which I rely on to identify their relative preferences for K-Cups brands.

3.3 Purchase Concentration in Brands

Figure 8.1 shows the Herfindahl-Hirschman Index (HHI, Herfindahl (1950) and Hirschman (1945)) of coffee purchases by the adopting households before their Keurig-machine adoptions and by the households who never adopted during the data period conditional on making at least five coffee purchases. Most of them (88.4%) have HHI above 2500, which is the Department of Justice's definition for high concentration (DOJ and FTC, 2010). Thus, households are not particularly variety seeking in the ground coffee category.[11] A mass of households (10.8%) has HHI of 10,000, which indicates they only purchase one brand for all of their purchases. I also don't observe strong differences between eventual adopters before their adoption and those never adopted. However, their HHI distributions are not completely overlapping, and adopters purchase little more variety.

Despite the slight apriori difference, households purchase more brands after their Keurig machine adoptions. Figure 8.2 shows the HHI of coffee purchases for the same households before and after their Keurig adoptions. The entire HHI distribution shifts to a much lower level after the adoption. Hence, after Keurig machine adoption, households tend to purchase more brands, though the relationship is not necessarily causal given potential selection bias. Appendix H presents results when I use concentration ratios instead of HHI. The distribution shifts are very similar.

11. The ground coffee category has a large number of brands (700+ brands purchased by homescan panelists), and thus the results is not driven by lack of variety in the ground coffee category.

Though the two distributions in figure 8.2 are over different time periods, the increase in brand purchase variety is not a simple time trend. To rule out the time effect, I create an artificial reference group and show no distributional shift for none Keurig-machine adoption households. By comparison with non-adoption households, I show the adopters purchase a greater coffee brand variety during the same period. To create the artificial reference group, I select the households that never adopted the Keurig machine, and compare their coffee purchase HHI in 2008 and 2009 with that in in 2012 and 2013. For comparison, I also choose the adopters with adoption date sometime in 2010 to 2011 and compare their coffee-purchase HHI in 2008 and 2009 with that in 2012 and 2013. Figure 8.3 shows the comparisons. The HHI distribution did not shift over time for those households that never adopted, and on the hand, the adopters experienced the shift to lower HHI before and after adoption.[12]

So, what causes the shift in HHI distribution for Keurig machine adopters? Are households buying more brands on a given trip, or are households switching more often between shopping trips? To answer these questions, I create summary statistics for both between-trip switching behavior and within-trip variety-seeking behavior (purchase of multiple brands). Figure 8.4 shows the percentage of trips with purchases of 1, 2,...,5 brands, and we observe an increase in the purchase of multiple brands in the same shopping occasion. Households still mostly buy one brand (about 85%) for each trip after adoption, but the change is sizable compared with that of before adoption. Figure 8.5 shows the percentage of trips buying a subset of last trip's brands. After the Keurig machine adoptions, households seem to be switching more often (median percent of trips purchasing a subset of last trip's brands went from 0.50 to 0.31, a 40% decrease) between shopping trips. By contrast, the reference group has

12. Appendix I explores the learning explanation for increased purchase variety. As show in figure L.3, most of the distributional shift in purchasing more variety after adoption is preserved two years after the adoption date.

mainly stayed at the same level of between-trip switching (percent of trips purchasing a subset of last trip's brands went from 0.58 to 0.60 on the median). However, without a formal model, we cannot rule out the temporal switching is entirely due to changes in the shopping environment, such as price changes.

3.4 Concentration in Brands and Flavors

Roast and flavors are essential dimensions of product characteristics in the coffee market. Different roast styles and flavors produce different tastes and aromas even for the same coffee beans. In this section, I summarize household purchase patterns when the product is defined as a brand, flavor, and roast-style combination. Compared to brand concentration, product concentration is much lower. Figure 8.6 shows the product HHI distributions before and after the Keurig adoption. We observe similar distributional patterns compared to the brand HHI figure - the distribution looks similar for non-adopters and adopters before their adoptions, and adopters shift to less concentrated purchases after their adoptions. Product purchases are overall much less concentrated compared to brand purchases. However, one should be cautious about over-interpreting the graph. In my context, switching brands means switching products because of the definition of a product. Appendix H presents the results when I use concentration ratios instead of HHI. The distributional patterns are very similar.

Because brand switching automatically leads to product switching, brand switching can drive the distributional shifts in product purchases. Thus, the distributional shifts in product purchases are confounded with brand purchases. To isolate the two, I summarize the product purchase behavior conditional on brand choices along two dimensions: cross-sectional (given trip) and temporal. Cross-sectionally, we decompose trips into four categories: (i) discrete choice - purchasing one unit of a product; (ii) multiple quantities of the same product; (iii) multiple flavors (roasts) of the same brand; and (iv) multiple brands. Figure 8.7 shows the cross-sectional choice decomposition by adoption status. Single unit discrete choice is the most prevalent, but as many as 23.8%-36.5% of the trips are not single unit choice. The largest source of non-discrete choice before the adoption is quantity choice at about 18%. After the Keurig adoption, multiple-brand purchases become more prevalent, increasing from 3.9% to 15%. Somewhat surprisingly, we don't see big increases in purchases of multiple flavors within a brand after the adoption - only increasing from 4.6% to 5.3%. Therefore, multiple-brand purchases is the largest source of increase in non-discrete choice cross-sectionally for Keurig adopters.

To isolate the temporal flavor switching behavior from brand switching, I only keep trips the household purchased a subset of the last trip's brands. Hence, these trips only include temporal flavor switchings within brands. Figure 8.8 shows the distribution of percent of trips purchasing a subset of last trips's flavors. We observe little distributional shift before and after adoption. Indeed, the median only had a small decrease from 0.57 to 0.50. Therefore, most of the distribution change in Figure 8.6 is due to brand switching, whereas flavor switching within a brand is largely unchanged.

3.5 Purchase Behavior after Adoption

Ground coffee and K-Cups are still substitutes, because households sometime still purchase ground coffee after their Keurig-machine adoptions. This feature of the data allows for estimating the relative preferences of ground coffee and K-Cups, which would otherwise be difficult. To show this data feature empirically, I decompose post-adoption trips into purchases of ground, K-Cup, or both. Similarly, I also decompose purchase quantity and expenditure into ground and K-Cup purchases. To avoid the households that permanently switched back to ground coffee, I only include the trips between their first K-Cup purchase and their last K-Cup purchase. Thus, the shares will most likely be a conservative estimate of ground shares. Nevertheless, we still observe sizable ground purchases - roughly 1/4 of the trips. Figure 8.9 shows the shares by type. Ground constitutes a sizable portion of servings purchased, but its expenditure share tends to be small because of its low prices per-serving.

To better understand household valuation for K-Cups, I summarize their annual coffee expenditures before and after Keurig-machine adoptions. Results show households spend more on coffee annually after their Keurig machine adoptions. Figure 8.10 shows the spending on coffee before and after the machine adoptions for adopters as well as for the reference group. I see a dramatic increase in median coffee spending from \$64.33 to \$170.32, which is mainly due to the high prices of K-Cups. For the reference group, the median spending only increases from \$59.17 to \$71.36 over the same time periods. The coffee price increases over time drive the expenditure increases for the reference group. For adopters, the large spending increase indicates they value the option to consume K-Cups.

CHAPTER 4

MODEL

Households may derive no utility from the hardware (Keurig machines) if s/he cannot consume the software (coffee in the Keurig case). Given the interplay of the hardware market and the software market, I develop a two-stage model of household demand. Although the chronological order of the two stages is adoption and then coffee consumption, the reverse order is easier to understand because coffee preferences feeds into hardware adoptions. Therefore, I first present the model for household coffee purchases, and then the adoption model.

4.1 Household's Coffee Consumption

Household coffee purchases do not follow a fixed time interval, and the quantity purchased each time may differ. I therefore model the coffee purchase as a two-step process: purchase decision, and then product and quantity choice. The purchase decision is a simple logistic process as a function of inventory of the type of coffee.[1] Conditional on the purchase decision, a household may purchase multiple brands, and thus I model the purchase as a multiple discreteness process as in Hendel (1999), Kim et al. (2002), Dubé (2004), and Bhat (2008). My model of the household purchases closely resembles that of Kim et al. (2002) and Bhat (2008). Their model builds on household utility theory and allows for the joint modeling of product purchased and quantity purchased.[2]

1. The inventory is imputed based on purchase rate, and Appendix G describes the details of the imputation procedure.

2. Because the coffee is technically a storable good, a dynamic multiple discreteness model would be ideal. However, the computational complexity would make the overall model intractable. One can view my two-step process as an approximation to the underlying dynamic process. When the inventory does not interact with product availability and prices, the approximation is exact.

Conditional on a purchase, I still allow an outside option, the numeraire, which anchors the scale of the utility. The utility anchoring avoids extreme utility gains for consumers who rarely switch brand choices, and also produces more reasonable welfare estimates. Conditional on the number of inside products chosen, the numeraire doesn't affect the choices. However, my treatment of the outside option doesn't allow for more frequent purchases to create category expansion. Instead, I handle category expansion through the purchase decision model, which is incorporated into utility gain calculations. The decision on the outside option is primarily driven by data, because I cannot find a suitable outside option for the category. Appendix J shows the effects of adoption on caffeinated drink, drink, grocery and total expenditures. Keurig machine adoption doesn't seem to have an impact on these options. Intuitively, the model represents the problem of how to best allocate the coffee spending so that the marginal utility of an extra dollar spent on coffee is identical to the marginal utility of that dollar spent on the numeraire.

4.1.1 Brand and Quantity Choice

Conditional on a purchase occasion at a retailer r, the household makes the decision regarding which product[3] and what quantity to buy. Assume the household's utility follows the following translated form,

$$U(x_0, x_1, \ldots, x_K) = \psi_0 x_0 + \left\{ \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left[\left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho},$$

where x_0 is the outside good quantity, ψ_0 is the outside good quality index, x_k represents the quantity of good k, $\theta_k = \{\alpha_k, \gamma_k, \psi_k\}$ denotes the corresponding parameter

3. In this paper, a product is defined as a combination of brand and flavor.

vector, and ρ induces stronger competition of the inside good. To ensure positive utility (normal goods) and non-increasing marginal utility, I constrain $\psi_k > 0$, $\alpha_k \leq 1$, and $\rho \leq 1$. As in Kim et al. (2002) and Bhat (2008), ψ_k is the quality index and controls for the household's marginal utility for good k. Both α_k and γ_k control the rate of satiation from good k (i.e., curvature of the utility function) and are hard to separately identify in the data. Given the empirical challenge, I normalize $\gamma_k = 1$ for all k as in Kim et al. (2002). Also, working with the expenditures instead of quantities is a bit easier.[4] Thus, I translate the utility expression to

$$U(e_0, e_1, \ldots, e_K) = \psi_0 e_0 + \left\{ \sum_{k=1}^{K} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho},$$

where the price of the outside good is normalized to 1. Let E be the total budget for coffee and the outside good, and the budget constraint can be expressed as[5]

$$\sum_{k=1}^{K} e_k = E.$$

To capture certain random nature of household choices, I allow the quality index ψ_k to include a random-shock term ε_k, which is observed to the household but not the econometrician, in the following form:

$$\psi_k = \psi(z_k, \varepsilon_k) = \psi(z_k) \cdot e^{\varepsilon_k},$$

4. This is without loss of generality as quantity demand $x_k = e_k/p_k$.

5. Here I assume firms charge a uniform price in the relevant region of household choice, and households are allowed to choose a continuous quantity of coffee. While seemly restricting, my research question is primarily focused on examining the role of variety, not quantity discount on household choice on two-sided platforms. In addition, I previously estimate a model with discrete choice with discrete quantity as Allenby et al. (2004), and results also indicate high K-Cup valuation and negative adjustment to state dependence as using the current model.

where z_ks are product characteristics. For analytical convenience, I normalize

$$\psi_0 = \exp(\rho\varepsilon_0),$$

for the numeraire. For other products, I assume linearity in exponents of the $\psi(\cdot)$ function,

$$\psi(z_k, \varepsilon_k) = \exp(z_k^{'}\beta + \varepsilon_k).$$

Specifically, I assume the quality index ψ_k for product k can be expressed as,

$$\begin{aligned}
\ln \psi_k &= \nu_k + \varepsilon_k \\
&= \underbrace{\zeta_k}_{\text{brand intercept}} + \underbrace{b_1 \mathbb{I}\{k \in \mathcal{K}\}}_{\text{K-Cup Fixed Effect}} + \underbrace{b_2 \mathbb{I}\{k \in s\}}_{\text{State Dependence}} \\
&\quad + \underbrace{b_3 \mathbb{I}\{k \in s \cap k \in \mathcal{K}\}}_{\text{K-Cup Adjustment}} + b_4 \ln N_k + \underbrace{b_5 L_k}_{\text{Flavor Effects}} + \underbrace{\varepsilon_k}_{\text{Latent Utility Draw}}
\end{aligned}$$

where $\mathcal{K}$ is the set of K-Cup brands, s is the set of last purchased brands, b_2 is the state-dependence parameter for ground coffee, b_3 is the state-dependence adjustment for K-Cups, b_4 is parameter for the Ackerberg and Rysman (2005) adjustment to approximate crowded product space, N_k is the number of brands in ground coffee or K-Cups, and ε_k is the type I extreme value latent utility term. The b_3 parameter is of particular interest since it indicates whether households are more likely to switch brands after their Keurig-machine adoptions. The b_4 parameter is also of great interest since it approximate the crowdedness of the product space and thus adding a brand will not always lead to an increase in the overall utility of the consumer. Given the set up, the household then solve the utility maximization problem to make choices.

4.1.2 The Purchase Decision

Households make coffee purchases at irregular intervals and may visit different stores to purchase coffee. Let t denote time, and the household may decide either to purchase coffee or not depending on the current coffee inventory level. Assume the probability of making a purchase follow the logistic process,

$$\Pr(y = 1) = \varphi_t(\iota) = \frac{\exp(\lambda_1 + \lambda_2 \mathbb{I}\{\iota = 1\} + \lambda_3 I_t)}{1 + \exp(\lambda_1 + \lambda_2 \mathbb{I}\{\iota = 1\} + \lambda_3 I_t)}.$$

Note observable heterogeneity can be incorporated into the model by treating λ as a function of observable characteristics. By incorporating household characteristics, I can predict the coffee purchase probability after the Keurig-machine adoption based on similar households, even if the household haven't adopted the Keurig machine yet.

Different retailers may have varied coffee selections. To accurately estimate the value of Keurig machine adoption, I need to take the varied coffee selections into consideration. The household may also visit different stores depending on their store-visit habits. In addition, the visit habit may shift due to the Keurig machine adoption. A full fledged store choice model is well beyond the scope of this paper. Instead, I assume the household visits a store r with probability given by

$$\Pr(r) = \phi_{rt}(\iota) = \frac{\exp(\alpha_r)}{\sum_{s \in \mathcal{R}} \exp(\alpha_s)}$$

where αs are parameters, and $\mathcal{R}$ is the set of retailers in the choice set of the household.

4.1.3 The Expected Gain

The household makes the Keurig-machine adoption decision based on potential gains in utility. As such, I characterize the expected flow utility gain from the Keurig machine adoption, which is the option value from consuming K-Cups. Let $\mathcal{G}_r$ and $\mathcal{K}_r$ denote the available ground coffee brand and K-Cup brands available, respectively. Conditional on a purchase of coffee at retailer r, the indirect utility function when both ground coffee and K-Cup coffee are available is

$$\mathbf{V}^*(G_r \cup \mathcal{K}_r) = \max_{\{e_0, e_k\}, k \in G_r \cup \mathcal{K}_r} \psi_0 e_0 + \left\{ \sum_{k \in G_r \cup \mathcal{K}_r} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho},$$

and when only ground coffee is available, the indirect utility function is

$$\mathbf{V}^*(G_r) = \max_{\{e_0, e_k\}, k \in G_r} \psi_0 e_0 + \left\{ \sum_{k \in G_r} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho}.$$

I assume Keurig machine adoption decision happens before the realization of ε for coffee consumption because the machine and K-Cups are often sold at different retailers. For example, the largest Keurig-machine retailers include Bed Bath & Beyond and Kohls, whereas the largest K-Cup retailers are often grocery stores. This feature is very different from that of Hartmann and Nair (2010), where razors and blades are often sold in the same retailer. As such, I assume the expected utility gain is prior to the realization of ε,

$$\delta_r = \mathbb{E}_{\varepsilon} \left[\mathbf{V}^*(G_{rt} \cup \mathcal{K}_{rt}) \right] - \mathbb{E}_{\varepsilon} \left[\mathbf{V}^*(G_{rt}) \right].$$

The above δ_r is conditional on making a purchase at a retailer r. Given the probability of making a purchase and shopping a particular retailer, the unconditional expected utility gain δ_t (adding the t subscript back) is,

$$\delta_t = \varphi_t(1) \cdot \sum_{r \in R} \phi_{rt} \mathbb{E}_{\varepsilon} \left[\mathbf{V}^*(G_{rt} \cup \mathcal{K}_{rt}) \right] - \varphi_t(0) \cdot \sum_{r \in R} \phi_{rt} \mathbb{E}_{\varepsilon} \left[\mathbf{V}^*(G_{rt}) \right].$$

The household then act on δ_t together with Keurig machine prices to determine whether to adopt the Keurig machine. Because φ is a function of inventory I and the expected maximum of utility is a function of last purchased brand s, price vector of coffee p, available ground coffee $\mathcal{G}$, and available K-Cup coffee $\mathcal{K}$, δ is a function of $\{I, s, p, \mathcal{G}, \mathcal{K}\}$. Let

$$x = \{I, s, p, \mathcal{G}, \mathcal{K}\}$$

denote the set of variables affecting δ.

4.2 Household Adoption of the Machine

Imagining a consumer makes the machine adoption decision based one period utility gain would hardly be reasonable because the Keurig machine is a durable good priced at \$100-\$200. Thus, my model of the Keurig-machine adoption assumes both the current and future utility gains from the K-Cups affect the adoption. To formulate the adoption model, I need to take a stance on whether consumers have rational expectations about future utility gains. If consumers do have rational expectations, they would consider the evolution of the future utility gain in their machine adoption decisions. However, if they don't have rational expectations, they may regard the future utility gains and the current utility gain to be the same. Figure 8.11 shows the estimated expected utility gain, and the increasing trend is rather clear. Because of this increasing trend over time, I assume consumers have rational expectations over the evolution of utility gains.

Based on figure 8.12, prices of Keurig machines follow the base price and promotion price patterns. As a result of this pricing pattern, strategic waiting by consumers could generate significant savings often in the magnitude of \$20-\$50. Moreover, few consumers purchase the Keurig machine at the regular price based on the purchase data and imputed prices. This purchase feature further make the strategic waiting conjecture plausible. Therefore, I assume consumers are strategic about the prices of the Keurig machine.

Because of the assumptions of rational exceptions over the future utility gains and strategic waiting, I model the adoption decision as a dynamic discrete choice problem. At the start of period t, I assume the household makes the adoption decision before the store trip for coffee purchase.[6] Conditional on the household not having adopted the Keurig Machine yet, let the flow value of adoption be denoted by

$$v_t(\iota_t = 1|\iota_{t-1} = 0, x_t, P_t, \eta_{1t}) \;=\; \tau - \vartheta P_t + \eta_{1t} + \underbrace{\mathbb{E}_{y,r}\left[\mathbb{E}_{\varepsilon}\left[\mathbf{V}^*(G_{rt} \cup \mathcal{K}_{rt})\right]|\iota = 1\right] - \mathbb{E}_{y,r}\left[\mathbb{E}_{\varepsilon}\left[\mathbf{V}^*(G_{rt})\right]|\iota = 0\right]}_{\text{Utility Gain } \delta_t}$$

where ι denotes the adoption status, τ is one-time hardware adoption utility, ϑ is the price coefficient for the hardware, P is the price index of the Keurig machines, and η_{1t} is a random utility shock.[7]The utility gain's inner part denotes the utility gain conditional on making a purchase $y = 1$ and visiting a retailer r, and outer expectation gives the unconditional expected utility gain. If the household don't

6. If the household makes the machine adoption decision in the same trip as coffee purchase, the current period utility gain would be different because of the realization of random shocks in coffee. However, given the time period is defined at the week level, and the household still has to make the adoption decision without observing the future random shocks, the loss due to the timing assumption should be small.

7. The price index is estimated by aggregating the prices of different Keurig machine models. Several models of Keurig machines are available, but data for each specific model are scant. The primary purpose of the paper is to estimate the effects of brand variety and quality of K-Cups on Keurig-machine adoption, and therefore, as long as the utility gain is well estimated and the general correlation between hardware price and brand variety and quality of K-Cups are included, I can estimate the effect of utility gain on machine adoption.

adopt the hardware, the household receive no extra gain:

$$v_{0t}(\iota_t = 0|\iota_{t-1} = 0, x_t, P_t, \eta_{1t}) = \eta_{0t},$$

where η_{0t} is a random utility shock. Once the household adopt the machine, I assume the household have the machine forever, and do not model the replacement decision. Therefore, if the household currently hold the machine, they receive the flow utility gain,

$$v_t(\iota_{t-1} = 1|x_t) = \delta_t(x_t).$$

Dropping the t subscript and using $'$ to indicate the next period state variables, I can write the Bellman equation as

$$\begin{aligned} V(\iota = 0, P, x, \eta) \quad = \quad \max \Bigg\{ & \beta \mathbb{E}\left[V(0, P', x', \eta')\right] + \eta_0, \\ & \tau - \vartheta P + \delta + \beta \mathbb{E}\left[V(1, x')\right] + \eta_1 \Bigg\}, \end{aligned}$$

where β is the discount factor. The number of brands in ground coffee and K-Cups amounts to hundreds, and solving a dynamic programming problem with such a high dimensional state space is virtually impossible. In this case, however, the set of state variables $x = \{I, s, p, \mathcal{G}, \mathcal{K}\}$ only affect the adoption utility through the flow utility gain δ. In other words, δ is a sufficient statistic for the set of states $\{I, s, p, \mathcal{J}, \mathcal{K}\}$. One can interpret δ as an inclusive value perceived by a bounded rational consumer because the consumer cannot conceivably handle such a high dimensional dynamic program. With this assumption, I can reduce the state space, and replace x with δ in the spirit of Melnikov (2013). After the dimensionality reduction, the Bellman equation simplifies to,

$$\begin{aligned}\mathcal{V}(\iota=0,P,\delta,\eta) &= \max\Bigg\{\beta\mathbb{E}_{\{P',\delta',\eta'\}}\left[\mathcal{V}(0,P',\delta',\eta')\right]+\eta_0,\\ &\quad \tau-\vartheta P+\underbrace{\delta+\beta\mathbb{E}_{\delta'}\left[\mathcal{V}(1,\delta')\right]}_{\mathcal{V}(1,\delta)}+\eta_1\Bigg\},\end{aligned}$$

because the value of holding the machine only depends on the per-period flow utility gain after adoption. Assuming the variance of η is not harmless, because I have normalized the utility in the coffee-purchase model. Therefore, I instead assume $\frac{\eta}{\sigma}$ follows the standard type I extreme value distribution with mean 0, and estimate σ in the model. Based on Rust (1987), the expected value function over η in closed form is,

$$\begin{aligned}\mathcal{W}(\iota=0,P,\delta) &= \int\mathcal{V}(\iota=0,P,\delta,\eta)dF(\eta)\\ &= \sigma\ln\left[\exp\left(\frac{\beta\mathbb{E}_{\{P',\delta',\eta'\}}\left[\mathcal{V}(0,P',\delta',\eta')\right]}{\sigma}\right)+\exp\left(\frac{\tau-\vartheta P+\mathcal{V}(1,\delta)}{\sigma}\right)\right]\\ &= \sigma\ln\left[\exp\left(\frac{\beta\mathbb{E}_{\{P',\delta'\}}\left[\mathcal{W}(0,P',\delta')\right]}{\sigma}\right)+\exp\left(\frac{\tau-\vartheta P+\mathcal{V}(1,\delta)}{\sigma}\right)\right].\end{aligned}$$

In estimation, the dimensionality reduction significantly simplies the estimation problem, because instead of solving the original dynamic program, I can solve the Bellman equation for $\mathcal{W}(\cdot)$.

State and State Transitions Specifying the evolution processes of the state variables P and δ completes the adoption model specification. To better capture the price process, I introduce another state variable the average price level $\bar{P}$, and assume it evolves deterministically as

$$\bar{P}_{t+1} = (1 - \omega)\bar{P}_t + \omega P_t.$$

The price in period $t + 1$ then follows a normal distribution around the mean,

$$\ln P_{t+1} \sim N(\mu(\bar{P}_{t+1}), \sigma_P^2).$$

In the context of this model,

$$\mu(\bar{P}_{t+1}) = \gamma_1 + \gamma_2 \ln \bar{P}_{t+1}.$$

The decision to introduce such a state variable and use this particular Markov process is driven by data. The Keurig machine prices don't follow the typical durable good pricing path, which is stochastically declining with time. Figure 8.12 shows the median price and sales volume of Keurig machine Series 3 in a large merchandiser chain in New York. We don't see a stochastically declining trend in price; rather, we observe the typical regular and promotion pricing. The regular price actually increased over time, and finally declined in 2013. The first vertical dashed line corresponds to the announcement date that the Folgers brand would be available on the Keurig platform, and the second vertical dashed line corresponds to the announcement date that the Starbucks brand would be available on the Keurig platform. The regular price increases happened roughly immediately before the announcements.

For the utility gain, I also assume δ follows the first-order Markov process, and

$$\ln(\delta_{t+1} + 1) \sim (\mu(\delta_t), \sigma_\delta^2).$$

In the context of this model,

$$\mathbb{E}\left[\ln\left(\delta_{t+1}+1\right)\right]=\mu(\delta_t)=\gamma_3+\gamma_4\ln\left(\delta_t+1\right).$$

Because I observe prices in the data, and estimate δ using the coffee purchase data, I can estimate the parameters $\theta^s=\{\rho,\gamma_1,\gamma_2,\sigma_P^2,\gamma_3,\gamma_4,\sigma_\delta^2\}$ from the observed states. This procedure avoids iteratively estimating the state parameters in the Bellman equation.

4.3 Household Heterogeneity

In this paper, I assume households are heterogeneous in their coffee preferences and thus adoption value, but homogeneous in other parameters of the machine adoption. Having heterogeneity in both the adoption problem and the coffee preferences is desirable. However, each household only adopt the Keurig machine once. Moreover, the Nielsen Homescan panel is not a perpetual panel with fixed households. Instead, some incumbent households may drop out, and new households may enter in any given panel year. At individual household level, the adoption decision is an optimal stopping problem. Thus, mechanically, a newly entered adoption household would appear to have a higher valuation over the platform compared to a household who has stayed in the panel for a long time because of a relatively shorter waiting time. One can solve this issue by constaining all households to have the same start and end period for the Keurig adoption. To do so, one must either drop a large quantity of data to have only comparable households or impute additional shopping environment data when the households are completely unobserved. I consider neither choices to be good options, and thus assume homogeneity in adoption model parameters. However, the adoption model is not homogeneous, because I already incorporate the heterogeneous preferences for coffee in the inclusive value δ_i, which is a large part of the consumer heterogeneity.

CHAPTER 5

ESTIMATION

I estimate the parameters in the two separate Bayesian models: (a) household preferences for coffee and per-period utility gain, and (b) conditional on per-period utility-gain estimates, parameters governing the adoption problem. The models should ideally be estimated in one algorithm using Gibbs sampling. However, such estimation is computationally infeasible because I have to estimate the flow utility gains and the Markov process parameters governing them for each Markov Chain Monte Carlo (MCMC) draw. Because the utility gain depends on the expected indirect utility function from coffee purchases, which has no closed-form, it's computed via Monte Carlo integration using random draws of latent utility shocks. In each Monte Carlo draw of the latent quality shock, I need to solve the utility-maximization problem for each household and time period (more than 10 million optimization problems for each Monte Carlo draw). Therefore, computing the utility gain is very computationally intensive, and I have to estimate the coffee consumption preferences and adoption parameters in two different MCMC chains to make the estimation feasible.

5.1 Coffee Preferences and Utility Gains

In the Homescan data, households purchased over 700 different ground coffee brands and over 50 of K-Cup brands. The top 10 brands including private label brands have 81% household spending share in ground coffee and 76% spending share in K-Cups. Most brands have only a few purchases. Estimating a brand intercept (ζ) for each brand with heterogeneous consumer preferences is nearly impossible. Instead, I estimate 13 separate brand intercepts including K-Cup brands for top brands. These brands have 68% expenditure share of K-Cups and 61% expenditure of ground coffee.

All private label brands are constrained to have one brand intercept, and the rest of the small brands are constrained to have one brand intercept. Hence, I estimate 15 brand intercepts in total. I supplement the brand intercepts with additional product characteristics to better capture consumer preference for coffee products. Additional product characteristics include the following variables:

- Dummy for flavored coffee
- Dummies for roast levels: light roast, medium roast, medium dark roast, and dark roast. Medium roast is the most common and treated as the base.
- Dummy for assorted varieties
- Dummies for special coffee or blend: Kona blend, Colombian coffee, and Sumatra coffee.
- Dummy for the wholebean form.

Given the number of brands, estimating a separate satiation parameter α for each brand is difficult. Instead, I model satiation parameter at coffee type level. All ground coffee products have a common satiation parameter α_1 , and all K-Cup products have a separate satiation parameter α_2. Adding back household heterogeneity in the coffee consumption, let $\theta_{hc} = [\zeta_h, b_h, \tilde{\alpha}_{1h}, \tilde{\alpha}_{2h}, \tilde{\rho}]$, where

$$\begin{aligned}
\alpha_{1h} &= \frac{\exp(\tilde{\alpha}_{1h})}{1+\exp(\tilde{\alpha}_{1h})} \\
\alpha_{2h} &= \frac{\exp(\tilde{\alpha}_{2h})}{1+\exp(\tilde{\alpha}_{2h})} \\
\rho_h &= \frac{\exp(\tilde{\rho}_h)}{1+\exp(\tilde{\rho}_h)}.
\end{aligned}$$

These formulations guarantee the satiation parameters of ground coffee α_{1h} and K-Cups α_{2h}, and the overall control parameter ρ_h are between 0 and 1.

Without loss of generality, I reorder the products so that the first $1, \ldots, M_{ht}$ products are the chosen ones at choice occasion t, and the rest are not chosen. Let e^*_{hkt} denote the expenditure on product k at choice occasion t. The individual likelihood for coffee purchases given parameter vector θ_{hc} is then,

$$
\begin{aligned}
CL_h(\theta_{hc}) \;&=\; \prod_{t=1}^{T_{ht}} P_{hjt} = \prod_{t=1}^{T_{ht}} |J_t(\theta_{hc})| \cdot M_{ht}! \cdot \\
&\quad \frac{\exp\left(\frac{(1-\rho_h)}{\rho_h} \ln \Omega_t(\theta_{hc}) - \frac{\ln \rho_h}{\rho_h}\right) \Pi_{j=1}^{M_{ht}} \exp\left(V_{jt}(\theta_{hc})\right)}{\left[\exp\left(\frac{(1-\rho_h)}{\rho_h} \ln \Omega_t(\theta_{hc}) - \frac{\ln \rho_h}{\rho_h}\right) + \sum_{k=1}^{K_{ht}} \exp\left(V_{kt}(\theta_{hc})\right)\right]^{M_{ht}+1}}
\end{aligned}
$$

where $V_{jht} = z_j\beta + (\alpha_{1h}\mathbb{I}\{j \in \mathcal{G}\} + \alpha_{2h}\mathbb{I}\{j \in \mathcal{K}\} - 1) \ln\left(\frac{e^*_{jht}}{p_{jt}} + 1\right) - \ln p_{jt}$, T_{ht} is the number of purchase occasions by household h, $J(\cdot)$ is the Jacobian for the change of variables in the integration, and

$$
\Omega_t(\theta_{hc}) = \sum_{k=1}^{M_{ht}} \frac{\exp\left(z_k\beta - V_k(\theta_{hc})\right)}{\alpha_{1h}\mathbb{I}\{j \in \mathcal{G}\} + \alpha_{2h}\mathbb{I}\{j \in \mathcal{K}\}} \left[\left(\frac{e^*_{hkt}}{p_{kt}} + 1\right)^{\alpha_{1h}\mathbb{I}\{j\in\mathcal{G}\}+\alpha_{2h}\mathbb{I}\{j\in\mathcal{K}\}} - 1\right].
$$

Appendix A derives the likelihood and Jacobian in detail.

To mitigate the potential sample selection, I incorporate observable heterogeneity in consumer preference estimation. The potential selection problem may arise if households with higher valuations of the Keurig platform are more likely to be observed in the adoption sample compared to the households with lower valuations. The non-adoption households are never observed to make a K-Cup purchase, and the K-Cups cannot be in their choice set. A naive hierarchical Bayesian model would use the behavior of the adopted households to infer the preferences of the non-adoption households subject to the covariance structure of the preferences. To the extent that ground coffee preferences are completely informative of K-Cup preferences, the naive estimation would not yield biased estimates. However, this relationship is unlikely. Therefore, I include household characteristics in modeling consumer heterogeneity. The characteristics include income levels, household size, house types, employment status, presence of children, race, annual total spending, and the concentration of their spendings. To the extent these characteristics and ground coffee preferences are informative about their K-Cup preferences, my preference estimates will not be subject to the sample selection issue. To guard against non-ignorable selection, I can use a sample selection correction technique, however, such techniques currently don't exist in the multivariate setting, where all the preference parameters for K-Cup brands together with any K-Cup specific parameters are unobserved for non-adoption households. Therefore, I try include as many observable characteristics as possible to mitigate any potential selection issues.

Following the Bayesian hierarchical framework in Rossi et al. (2005), assume multivariate normal heterogeneity in parameters,

$$\theta_{hc} \sim MVN(\Delta' Z_h, \Sigma),$$

where Z_h is the vector of observed characteristics. Given this set up, the model can be estimated using Gibbs and Metropolis-Hastings (MH) algorithms. Following Rossi et al. (2005), let the prior be of the following distributions,

$$\begin{aligned} \Sigma &\sim IW(\nu_0, V_0) \\ \mathrm{vec}(\Delta) &\sim MVN(\mathrm{vec}(\bar{\Delta}), \Sigma \otimes A^{-1}), \end{aligned}$$

and the prior parameters are,

$$
\begin{aligned}
\nu_0 &= \max(4, 0.01n) \\
s_l &= \frac{1}{n-1}\sum_h^n (z_{hl} - \bar{z}_l)^2 \quad \text{variance of } z_l \\
S_Z &= \operatorname{diag}(s_1^2, ..., s_L^2) \quad \text{stacked diagonal variance} \\
A &= \nu_0 S_Z \\
V_0 &= \nu_0 I_m.
\end{aligned}
$$

where n is the number of number of households, and m is the dimensionality the individual parameters. The conditional posterior distributions are

$$
\begin{aligned}
p(\Sigma|\theta_{1c}, \dots, \theta_{Nc}) &\propto L(\Sigma|\theta_{1c}, \dots, \theta_{Nc}) \cdot p(\Sigma) \\
p(\Delta|\theta_{1c}, \dots, \theta_{Nc}, \Sigma) &\propto L(\Delta_c|\theta_{1c}, \dots, \theta_{Nc}, \Sigma) \cdot p(\Delta) \\
p(\theta_{hc}|\Delta, Z_h, \Sigma, D_h) &\propto CL(\theta_{hc}|D_h) \cdot p(\theta_{hc}|\theta_c, Z_h, \Sigma),
\end{aligned}
$$

where $L(\cdot)$ denote the likelihood function and D_h denote the relevant data for household h. By prior assumptions, the priors and posteriors of Σ and Δ are conditionally conjugate, which allows for direct drawing from the posterior. Because the prior and posterior distributions of individual specific parameters are not conjugate, I use a random walk MH algorithm to draw from the posterior. The posterior means of individual preferences are used to compute the value of having the option of K-Cups for each household and time period.

The final estimation data have 34,249 households, 680,261 purchase occasions and a total of 28 million observations including data for alternatives. The magnitude of the data coupled with a rather complex likelihood function create a computational challenge. To deal with the challenge, I estimate the model using parallel computation on a cluster. Following Rossi et al. (2005) suggestion, I parallelise the MH step for individual specific draw, which greatly reduce the computational time. My parallization strategy works in this setting because the individual specific draws are computational costly, and the communication costs are small given the small number of households. To improve the efficiency of the MH step, I use a random walk MH algorithm with step size tuned based on the fractional likelihood method in Rossi et al. (2005). Appendix B.1 details the posterior and algorithm steps.

The utility gain calculation The multiple discreteness model has no closed form indirect utility function. The model assumes households make their machine adoption decisions before the realization of ε, and thus the computation of utility gain integrates over ε. Without a closed form indirect utility, I use the Monte Carlo integration technique to obtain the expected utility gain. For each Monte Carlo draw of ε, I will need to solve the utility maximization problem for each household and potential shopping occasion, which amount to more than 10 millions constrained optimizations in each draw. The computational burden calls for an efficient optimization algorithm specific to this problem. Appendix C presents a fast algorithm to compute the maximized utility and corresponding expenditures via some simple algebra and one root calculation.

For each household h, potential retailer r, and time t, I compute the expected indirect utility of having the option of both ground coffee and K-Cups as,

$$\mathbb{E}_{\varepsilon}\left[\mathbf{V}_h^*(G_{rt} \cup \mathcal{K}_{rt})\right] = \\ \frac{1}{D}\sum_{d=1}^{D}\left[\max_{\{e_{hkt}\},k\in G_{rt}\cup\mathcal{K}_{rt}} U_h\left(E - \left(\sum_{k\in G_{rt}\cup K_{rt}} e_k\right), e_1, \ldots, e_{|G_{rt}\cup K_{rt}|}\right)\middle|\varepsilon_d\right],$$

and when only ground coffee is available, the expected indirect utility is,

$$\mathbb{E}_{\varepsilon}\left[\mathbf{V}_h^*(G_{rt})\right] = \frac{1}{D}\sum_{d=1}^{D}\left[\max_{\{e_{hkt}\},k\in G_{rt}} U_h\left(E-\left(\sum_{k\in G_{rt}} e_k\right), e_1,\ldots,e_{|G_{rt}\cup K_{rt}|}\right)\middle|\varepsilon_d\right].$$

The purchase incidence model is estimated using a simple logistic regression. For non-adoption households, the parameters allowing for differential purchase rate before and after adoption is predicted using the rest of the parameters together with household demographics. Store visit probabilities are estimated using simple shares of visits. With these probabilities, the household h's utility gain at time period t is,

$$\delta_{ht} = \varphi_{ht}(1)\cdot\sum_{r\in R}\phi_{hrt}(1)\mathbb{E}_{\varepsilon}\left[\mathbf{V}_h^*(G_{rt}\cup\mathcal{K}_{rt})\right] - \varphi_{ht}(0)\cdot\sum_{r\in R}\phi_{hrt}(0)\mathbb{E}_{\varepsilon}\left[\mathbf{V}_h^*(G_{rt})\right].$$

The estimated expected utility are then used to estimate the state evolution process and the machine adoption parameters.

5.2 Machine Adoption

For given parameters $\theta_a = \{\tau, \vartheta, \sigma\}$, I can solve the expected value function $\mathcal{W}(\iota = 0, P, \delta)$. Hence, choice-specific value functions are

$$\begin{cases} v_0(\iota = 0, P, \delta|\theta_a) = \beta\mathbb{E}_{\{P',\zeta'\}}\left[\mathcal{W}(\iota = 0, P', \delta'|\theta_a)\right] & \text{if not adopting} \\ v_1(\iota = 0, P, \delta|\theta_a) = \tau - \vartheta P + \mathcal{V}(1, \delta') & \text{if adopting} \end{cases},$$

where β is set to be 0.995 at weekly level in estimation because it is not identified. Recall the random utility shock with a normalization parameter, $\frac{\eta}{\sigma}$ follows a standard type I extreme value distribution with mean 0. Give these settings, the conditional adoption probability is

$$\Pr(y=1|\iota=0,P,\delta)=\frac{\exp\left(\frac{v_1(\iota=0,P,\delta|\theta_a)}{\sigma}\right)}{\exp\left(\frac{v_0(\iota=0,P,\delta|\theta_a)}{\sigma}\right)+\exp\left(\frac{v_1(\iota=0,P,\delta|\theta_a)}{\sigma}\right)},$$

where y is the action chosen, and 1 means the Keurig-machine adoption.

Figure 8.12 shows the sales of Keurig machines have strong seasonality. To control for seasonality, I treat seasonality as shocks to households, and the modified adoption probability is,

$$\Pr(y=1|\iota=0,P,\delta)=\frac{\exp\left(\frac{v_1(\iota=0,P,\delta|\theta_a)}{\sigma}+X_{ht}\kappa_2\right)}{\exp(\frac{v_0(\iota=0,P,\delta|\theta_a)}{\sigma})+\exp\left(\frac{v_1(\iota=0,P,\delta|\theta_a)}{\sigma}+X_{ht}\kappa_2\right)}.$$

where X_{ht} is the vector of seasonality and other utility shocks not anticipated by households. Though households may perfectly anticipate the coming of holidays, they may not fully anticipate the positive utility shock from the machine adoption. Also, I don't observe dips in sales before high-sales seasons, which otherwise should be the case given seasonality-strategic households. In addition, some households may get the Keurig machine as a gift, which they cannot anticipate. This assumption also greatly simplifies the estimation because it avoids including seasonality in the state space. Since seasonality doesn't coincide perfectly with the week, the transition of seasonality state variables may not be well defined.

Adding the individual subscript and time subscript back, the joint likelihood

$$\begin{aligned}L(\theta_a,\sigma,\kappa|a,P,\delta,X,F) &= \prod_{h=1}^{N}\prod_{t=1}^{T_h}(\Pr(y_{ht}=1|\iota_{ht}=0,P_{ht},\delta_{ht}))^{y_{ht}}\\ &\times\ (\Pr(y_{ht}=0|\iota_{ht}=0,P_{ht},\delta_{ht}))^{y_{ht}},\end{aligned}$$

where T_h is the time of adoption for household h. If household h has never adopted the machine, T_h is the last date observation of the household.

I use the Imai et al. (2009) (IJC) algorithm to estimate the dynamic adoption problem. The IJC algorithm offers the advantage of significantly reducing the computational burden by using the Bayesian method and value function approximation. In my context, I set the priors to be

$$[\theta_a, \kappa] \sim MVN(0, 100I_{||[\theta_a,\kappa]||}).$$

Section B.2 in the Appendix explains the detailed estimation procedure.

5.3 Identification

Given the model has two layers, I need to estimate the parameters associated with each layer of the model requiring specific variations in the data. The data on repeated purchases of coffee help identify the household preferences for different brands of coffee. The estimated parameters in preferences then feed into the machine-adoption problem. Together with the price variation of Keurig machines and the timing of adoption, they help identify parameters of the adoption model.

The household-preference parameters for coffee are identified through the repeated purchases of coffee and relative price variations during those purchases. If a consumer purchases a product repeatedly even if the product has a relatively higher price compared to other products, the household is identified as having higher preferences for the product. Price promotions may trigger a consumer to switch to a different brand, and if the consumer then still purchases the same brand after the promotional period, the household displays state dependence. The relative price variations identify the magnitude of each parameter. Finally, households stay in the panel for multiple periods and make multiple purchases, which identify how they differ (heterogeneity) from each by looking at their choice patterns.

The estimated household-preference parameters and the shopping environment allow me to back out the consumption values of coffee brands. The consumption values help identify the parameters of interests in the machine-adoption problem. The prices for machines vary over time, and they help identify parameters in the adoption model.

CHAPTER 6

ESTIMATION RESULTS

6.1 Coffee Consumption

Estimation results show both mechanisms of household heterogeneity and variety seeking are at play on the Keurig platform. Households have heterogeneous preferences for K-Cups as shown, and figure 8.13 shows the preference distribution for K-Cups. Households exhibit a lower satiation parameter (stronger satiation) toward K-Cups versus ground coffee, which increases the probability of households choosing multiple brands in a given trip. For temporal switching, the state-dependence parameter is much weaker for K-Cups compared with that of ground coffee, which can lead to more temporal switching. I also find households treat K-Cups brands as more homogeneous than ground coffee, which can result in the choice of multiple brands on a given trip or brand switching between trips. Morever, brand dummies can explain the variations in relative brand preferences across households, indicating a significant vertical component in household preferences for coffee. The preference difference between the ground and K-Cups is the principal driver for this vertical component.

Because the utility is anchored against the dollar, I can compute the relative dollar value of each product to measure consumer preference. Table 8.1 shows the mean valuation for one serving of coffee at weekly level over brand and flavors with a 95% posterior coverage region. It shows, on average, households have a much higher valuation for K-Cups than for ground coffee. We also see observe variations in mean brands valuations, but they are much smaller compared to the valuation difference between K-cups and ground coffee.

To better understand the distribution of preference for K-Cups (K-Cup fixed effect), figure 8.13 shows the posterior density distribution of households with 95% credible intervals. K-Cup preference is high, but as expected, substantial heterogeneity exists. The estimated distribution indicates K-Cups are of great value to the a large percentage of households.

Figure 8.14 shows the distributions of satiation parameters with 95% credible intervals. Households have nearly no satiation over ground coffee in the range of their purchase quantity, which helps explain 93% of the ground coffee purchases consist of one brand and flavor. By comparison, households have a relatively lower satiation parameter (high satiation) in K-Cups, which explains why only 79% of the K-Cup purchases are of one brand and flavor.[1] The product pack size may be a driver of this result. Households may want to consume a variety of ground coffee brands as well, but ground coffee pack size is generally large which prevents households from choosing multiple brand at the same time. For example, the typical small pack size for ground coffee is 20-24 servings (10OZ - 12OZ), and in comparison, K-Cups often come in 12-18 servings. In addition, ground coffee is harder to preserve once opened for consumption, whereas K-Cups are individually packed which allows for long lasting freshness, and thus make consuming a variety of coffee easier. The interpretation of the parameters would depend on the exact mechanisms for generating the high satiation parameter. However, the exact interpretation is less important to this paper, because the parameter is only to capture the preference for variety on a given choice occasion, which is enough for exploring the mechanism of cross-network effects.

The state-dependent parameter is also of great interest since strong state-dependence parameter could lead households to switch less often. For example, if a consumer

1. Figure L.5 in Appendix L shows the scatter plot of MCMC draws of the satiation parameters for the ground coffee and K-Cups.

chooses his/her favorite brands on the last trip, s/he is less likely to change to a different brand when state dependence is present. Thus, a strong state dependence serves to reduce temporal switching because the consumer is often in the state of purchasing his/her favorite brand. Figure 8.15 shows the posterior density plots with a 95% credible interval by ground coffee and K-Cup adjustment. The state dependence of K-Cups is the sum of the state-dependence parameter plus the K-Cup adjustment. Thus, the negative K-Cup adjustment parameter indicates households have lower state-dependence parameters for K-Cup brands compared to ground coffee brands, and it contributes to more temporal switching. In other words, the Keurig platform makes switching brands between choice occasions easier.[2]

To better understand the nature of preference variation, I demean coffee preferences at household level and stack all the brand-preference parameters into a vector and regress them on whether the brand is a K-Cup brand and brand identities. By demeaning, I measure the relative preferences for different brands at household level, which help understand relative preferences for different brands. Figure 8.16 shows the R-squared of the posterior draws by regression types. The K-Cup dummy explains a substantial amount of variation in preferences, and adding brands only slightly increases the fit. The results indicate the vertical differentiation of K-Cups compared to ground coffee is strong.

Households could treat all K-Cup brands as being more similar compared to ground coffee, which is may be a result of similar packing and joint market activities. If so, triggering households to switch brands would be easier compared to more differentiated brands. To explore this possibility, I compare the variance of brand preference for K-Cup brands and ground coffee brands. Figure 8.17 shows histograms

2. Figure L.6 in Appendix L shows the scatter plot of MCMC draws of mean state dependence and K-Cup adjustment.

of the household-level variance analysis. The results indicate the variance is much lower for K-Cup brands than for ground coffee, which implies consumers indeed treat K-Cup brands as more similar. This finding helps to explain the variety increase in purchases. However, the finding does not undermine my previous results, because I explicitly estimate brand preferences in the model.

6.2 Machine Adoption

In the current model, price interacts with other state variables such as the current utility gain and mean price level to determine the Keurig machine adoption. For example, if the price is lower, but higher than the reference price, the household may still opt to adopt the machine next period. Moreover, the state-transition process plays a vital role in determining the value function and ultimately adoption probability. I first estimate the state transition parameters outside of the dynamic program and then use the estimates in the adoption problem. Note that these parameters are re-estimated for counterfactual analysis because consumers are assumed to form rational expectations over the future utility gains.

To investigate the effect of prices and utility gain on adoption, I compute the adoption probability, controlling for the mean price level at \$140, about the average price of Keurig machines. Figure 8.18 shows the conditional choice probability at various levels of the Keurig-machine price index and the per-period utility gain δ. The result indicates the adoption probability increases with increasing δ and decreasing machine price. Moreover, the adoption probability is low if prices are high, even if the utility gain is high. This feature arises because the mean price level is held at \$140, and high prices may trigger households to wait for the next period. Such behavior is consistent with the observation that most households purchased their machines on promotion. In general, the estimated model can capture the adoption behavior.

CHAPTER 7

COUNTERFACTUALS

In this section, I simulate counterfactual scenarios to understand better the mechanisms through which K-Cup variety functions to increase hardware adoption, and the impact of third-party brands on hardware adoption, Keurig's own K-Cup revenue, and consumer welfare. To understand the mechanisms through which K-Cup variety generates Keurig machine adoptions, I first simulate to see what happens to Keurig machine adoption when all households can only choose their most preferred brands. Each households may have a different most preferred brand, which may vary over time. Hence, the simulation results shows how much K-Cup variety is of value to consumer adoptions through providing them an opportunity to find a brand they really like. Second, I simulate the Keurig machine adoption rate when all households have a library of K-Cups to choose from, but their most preferred brand is replaced with an average K-Cup brand. Thus, the simulation results shows how much K-Cup variety is of value to consumer adoptions by satisfying their demand for variety. Through the two counterfactual simulations, I explore the relative importance of each mechanism to Keurig machine adoption.

The classical open versus closed platforms is also of great interest. As such, I simulate the counterfactual scenario if Keurig were a closed platform. The particular interest is on counterfactual Keurig machine adoption, revenue of Keurig owned brands, and consumer welfare. First, I show the Keurig adoption rate would have been much lower if third-party brands were not available on the platform, and the effect size is moderated by the consumer preference distribution. For indirect network effects, I show that GMCR-owned brands revenue would have been much lower if were not for third-party brands. In welfare analysis, I show large welfare loss the third party brands were not available on the Keurig platforms.

Thoughout the counterfactual analysis, I hold the supply sided condition fixed. This assumption may not be as a strong as it appears given the history of the Keurig platform. According the Anderson (2005) case study, the Keurig company was a startup, and had to contract with several roasters to provide varieties. By contrast, the Tassimo platform owned by Kraft mainly had Kraft owned brands. GMCR acquired the Keurig company in 2006 and continued the legacy. The legacy consideration may be a driver of the decision allowing third parties to produce K-Cups. Thus, whether GMCR's decision to continue providing variety in K-Cups is entirely driven by rational profit maximization motive is unclear. Moreover, in my talk with an industry insider from Starbucks, GMCR's contracts with each roaster vary, and negotiations determine the final contract terms. Though the Starbucks employee cannot provide too much information, he told me that Starbucks had some pricing power, but GMCR helped to determine the final prices. For some roasters, GMCR prices and distributes their K-Cups. The specific fee structures are strictly confidential. Modeling the bilateral bargaining is well beyond the scope of this paper. As such, I hold supply conditions fixed.

7.1 Mechanisms of K-Cup Variety to Keurig Adoptions

Figure 8.19 shows the expected Keurig machine adoption rate by availability of K-Cups plotted over time. Keurig adoption rate would be about 1/4 of current level by the end of 2013 if households can only choose their most preferred brand. By comparison, the Keurig machine adoption rate would be about 2/3 of current level if households have a library to choose from, even if they cannot find their most preferred brand. Therefore, the most preferred brand is important to Keurig machine adoption, but a library of K-Cup brands is vital to the Keurig platform growth. The results show demand for variety is the main mechanism through which K-Cup variety generates Keurig machine adoption.

Practically, the Keurig platform could be launched in the home segment with Keurig-owned brands, licensed brands, and authorized third-party brands.[1] Therefore, GMCR would have to make the decision about whether to license other brands or to open up its platform to third-party roasters. The change of ownership to GMCR in 2006[2] and several acquisitions made precisely determining the nature of relationships between several K-Cup brands and GMCR difficult. Because the distribution channel was not well set up before 2006, I take the ownership structure in 2006 as given and estimate counterfactuals based on it. Hence, all brands owned by GMCR in 2006 are considered owned, and brands acquired later are not owned brands. Table 8.2 shows GMCR-owned brands, licensed, and third-party brands in 2006.

The introduction of Keurig 2.0 using the digital rights management system in the fall of 2014 sparked great controversy, and several coffee roasters, including TreeHouse food and Rogers Family, took GMCR to court. The third-party roasters claimed GMCR's unnecessary "lockout" for the expired patent was anti-competitive, and hurt consumer welfare. Because consumers view K-Cup brands as more similar to each other than for ground coffee, how much GMCR's action hurt consumer welfare is unclear. In light of this welfare concern, I examine the impact of variety on household welfare when prices stay at current levels. Assuming no price changes is a limitation of

1. Unauthorized third-party brands started to come into existence in 2012 when the K-Cup patent was set to expire in September of that year. They are relatively small in market share by the end of 2013.

2. GMCR was investing in Keurig and was a major shareholder before complete acquisition.

the model because GMCR repeatedly increased the K-Cup prices as the penetration rate increased.

7.2 Open versus Closed Platform

7.2.1 Adoption

Reducing the variety supplied on the Keurig platform would decrease the adoption probability through decreasing utility gains. Figure 8.20 shows the plot of the average adoption rate by K-Cup variety supplied. The adoption rate declines significantly if GMCR only has its owned brands. Both third-party brands and licensed brands add value and increase adoption of the Keurig platform. Both consumer demand for variety and heterogeneity could predict such evolution path of adoption rate.

7.2.2 Value of lost households

Although allowing third-party producers increases household adoption and thus increases customer base, GMCR may also have to face competition from them. Though GMCR can share the profits in the form of royalty fees and other associated fees, whether the participation of these brands increase the revenue of GMCR-owned brands is unclear. Figure 8.21 shows the expected revenue of all K-Cups by variety supplied. In this particular setting, we observe the overall revenue on the Keurig platform declines when third-party brands or licensed brands were not available, which is a direct result of the declining adoption rate.

Figure 8.22 shows the GMCR-owned brand expected revenue by K-cup variety supplied. In this particular setting, I observe GMCR revenue declines if the platform does not offer third-party brands. The elimination of third-party brands leads to a considerable reduction in household adoption, which overweighs the benefits from reduced competition. Thus, the net effect of the removal is negative on GMCR-owned brand revenue. Of course, third-party brand need to pay royalty fees, because GMCR held the patent for K-Cup before the patent expiration in September 2012. The fees make it unclear how much the overall profits increase when third party brands are allowed.

7.2.3 Welfare

Figure 8.23 shows the compensating value percentage by household and variety supplied conditional on adoption of the Keurig machine. In this analysis, we observe high estimates of compensating value if K-Cups were not available. The CV percentage median is at 49% (panel 1). When only GMCR owned brands were available on the Keurig platform (Panel 2), the median CV percentage is 33%. When only third-party brands were unavailable, the CV percentage is 28%. Of course, the model may underestimate of the CV since GMCR may have increased K-Cup prices if third party brands were not available in absence of competition in the aftermarket. Some estimates are indeed very high, which is due to extreme preferences by some households who never changed their choices. Given the limitations, the estimates should be interpreted with great caution, but nevertheless highlight the value of K-Cups and the significance of variety for household welfare. In addition, some households actually don't require compensation for reduced variety on the Keurig platform, which is the result of Ackerberg and Rysman (2005) adjustment. With crowded product space, increasing varieties available don't always increase consumer welfare. For example, a consumer who strongly prefers GMCR brand to any other brand may be better off when only GMCR brand is available since added variety only increases confusion and mental load of the consumer. In this particular setting, I find 21% of households have negative CV when only GMCR owned brands were available, and 22% of households have negative CV when both GMCR owned brands and licensed brands were available.

CHAPTER 8

DISCUSSION AND CONCLUSION

In this paper, I first show households purchase a greater variety of brands after their Keurig-machine adoptions. I then estimate a hierarchical Bayesian demand model with multiple discreteness allowing for multivariate normal preference heterogeneity. The estimation results indicate households have high satiation, lower state dependence, and high preference for K-Cups. The high satiation results in the choice of multiple brands in the same choice occasion, and lower state dependence allows households to switch more often. Moreover, a simple variance analysis reveals preferences for K-Cup brands are more homogeneous (smaller between-brand variance) compared with ground coffee brands. The lower level of differentiation in household preferences enable households to more easily switch brands in K-Cups than in ground coffee.

After estimating household preferences, I link their preferences to their adoption of the Keurig machine. Because the number of state variables governing the utility gain (option value of K-Cups) is very high, estimating a general model of the household adoption problem is not feasible. As such, I resorted to a dimensionality-reduction technique by deriving the sufficient statistic for the utility gain. This approach is in the spirit of Melnikov (2013) and Gowrisankaran and Rysman (2012), but my approach requires no assumption on the value function transition. Estimation results indicate household utility gains are positively related to the adoption of the Keurig machine, and few are predicted to purchase above the mean level of prices.

To get a sense of the importance of third party brands in determining adoption (cross-network) and revenue for GMCR-owned brands (indirect network effects), I conduct counterfactual analysis by varying the level of variety supplied. I find the Keurig platform wouldn't have had such a high adoption rate if third-party brands and licensed brands were not available. Moreover, GMCR-owned brand revenue would have been lower if third-party brands were not available, which is due to indirect network effects.

Although I chose to focus here on an appliance platform, this paper provides a framework for the analysis of other similar platforms. The Keurig platform is perhaps the most successful household appliance in the last decade, going from almost no penetration to 25%-30% US household penetration rate NCA (2016). The tied feature of Keurig machines makes it ideal for the study of two-sided platforms and associated network effects. The study of network effects requires us to estimate household preferences. In most durable goods setting, estimating preferences is difficult because a consumer often purchases each software only once. By contrast, the repeated purchase feature of coffee allows me to estimate household preferences with much weaker assumptions, and the variation in prices and availability makes the identification transparent. The dynamics in the Keurig platform are similar to many platforms. For example, Amazon, Barnes and Noble, Best Buy, and BHPhotoVideo, as well as many e-commerce websites, started with selling their own products, but some have eventually allowed third parties to sell on their platform. Tied physical goods sellers (e.g. cars and car parts) also need to determine whether they should allow third parties to manufacture and sell compatible aftermarket goods. The most popular single-serving coffee platform in Europe, Nespresso which can also make Espresso, had remained a closed platform for the last decade, and Nespresso coffee is the only available coffee brand for the machine. The ultimate decision on whether the platform should supply more variety - either through building a new brand or recruiting third-party sellers - depends on the target consumer's preference distribution.

Substantively, this paper contributes to a better understanding of the mechanisms through which variety generates cross-network effects on two-sided platforms and more generally tied goods. In addition, this paper provide some direct evidence of indirect network effects, from which GMCR-owned brands received better revenue because of the participation of third-party brands. Moreover, the product space may become too "crowded" and direct-competition effects dominate indirect-network effects at some point, from which the participation of third-party brands leads to a decline in the revenue of GMCR-owned brands. The paper also provide new micro-level evidence for cross-network effects, which is a direct result of individual behavior.

Methodological, this paper develops a framework to analyze micro-level tied good, where the aftermarket good is repeatedly purchased and the primary good is purchased once. The framework developed here allows firms to understand how brand variety increases a consumer's expected utility gain and thus platform adoption. The machinery is potentially generalizable to other two-sided platforms where individual-level data are available and consumer preferences for the software could be estimated. Moreover, the inclusive value approach proposed in this paper allows firms and researchers alike to solve the curse of dimensionality problem when the platform adoption decision is dynamic, thus making price-promotion optimization possible. In addition, I extend the Kim et al. (2002) and Bhat (2008) multiple discreteness model to allow for an outside while preserving stronger competition among inside goods. The model also has the advantage of anchoring the scale of the utility function, which limits the magnitude of compensating value in welfare analysis. Though similar to the original model, the indirect utility function has no closed form, I developed a fast algorithm to compute the maximized utility for any random utility shock, which requires only one univariate root calculation.

The paper also suffers from some limitations - most notably, the imputation of data. The retail price data for K-Cups are relatively easy to impute because most chains follow the chain-region-specific pricing (Hitsch et al., 2017). However, the hardware data are much harder to impute because retailers sell various Keurig machine models at different prices at any point in time, and only a handful of households make purchases in every period. Given such a limitation, I only impute one price index of hardware. With measurement error, the price index will attenuate the price coefficient for machine adoption. Another limitation is the imputation of the adoption date. Although some households report their purchase of Keurig machines, others are only observed to make K-Cup purchases. For them, I impute the adoption date as the first K-Cup purchase week subject to the burn-in period. Appendix F discusses the detailed imputation procedure for adoption date and Keurig machine index. Another limitation is the lack of a well-defined supply model.

In summary, both heterogeneity and variety contribute to cross-network effects, and variety-seeking behavior seems to be a stronger force in driving adoption compared with heterogeneity. The business implication depends on the degree of variety seeking and heterogeneity of the focal platform. For the Keurig platform, allowing third-party brands to enter the platform improves the adoption probability and increases revenue for Keurig-owned brands. However, if household preferences change as simulated in section 7, allowing third-party sellers might not be in the best interest of the Keurig platform, as in the case of non-variety-seeking households. Future work will focus on relaxing the supply restriction, and on the robustness to assumptions of imputation procedures.

REFERENCES

Ackerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *The RAND Journal of Economics 36*(4), 771–788.

Allenby, G. M., T. S. Shively, S. Yang, and M. J. Garratt (2004). A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. *Marketing Science 23*(1), 95–108.

Anderson, E. T. (2005). Keurig at home: Managing a new product launch. *Kellogg School of Management Cases*, 1–19.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics 37*(3), 668–691.

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

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), pp. 841–890.

Bhat, C. R. (2008). 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. A Tribute to the Career of Frank Koppelman.

Broniarczyk, S. M., W. D. Hoyer, and L. McAlister (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research 35*(2), 166–176.

Caillaud, B. and B. Jullien (2003). Chicken & egg: Competition among intermediation service providers. *The RAND Journal of Economics 34*(2), 309–328.

Chen, J., U. Doraszelski, and J. E. Harrington (2009). Avoiding market dominance: Product compatibility in markets with network effects. *The RAND Journal of Economics 40*(3), 455–485.

Chintagunta, P. K. (1998). Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science 17*(3), 253–270.

Chintagunta, P. K., M. S. Qin, and M. A. Vitorino (2016). Licensing and price competition in tied-goods markets: An application to the single-serve coffee system industry.

Clements, M. T. and H. Ohashi (2005). Indirect network effects and the product cycle: Video games in the u.s., 1994–2002. *The Journal of Industrial Economics 53*(4), 515–542.

DOJ and FTC (2010). Horizontal merger guidelines.

Dubé, J.-P. (2004). Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science 23*(1), 66–81.

Dubé, J.-P., G. Hitsch, and P. Chintagunta (2010). Tipping and concentration in markets with indirect network effects. *Marketing Science 29*(2), 216–249.

Farrell, J. and G. Saloner (1985). Standardization, compatibility, and innovation. *The RAND Journal of Economics 16*(1), 70–83.

Farrell, J. and G. Saloner (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *The American Economic Review 76*(5), 940–955.

Gandal, N., M. Kende, and R. Rob (2000). The dynamics of technological adoption in hardware/software systems: The case of compact disc players. *The RAND Journal of Economics 31*(1), 43–61.

GMCR (2009). Fiscal 2008 annual report. Technical report, Green Mountain Coffee Roasters Inc.

Gourville, J. T. and D. Soman (2005). Overchoice and assortment type: When and why variety backfires. *Marketing Science 24*(3), 382–395.

Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy 120*(6), 1173–1219.

Hartmann, W. R. and H. S. Nair (2010). Retail competition and the dynamics of demand for tied goods. *Marketing Science 29*(2), 366–386.

Hendel, I. (1999). Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies 66*(2), 423–446.

Herfindahl, O. C. (1950). Concentration in the steel industry.

Hirschman, A. O. (1945). *National power and the structure of foreign trade*, Volume 105. Univ of California Press.

Hitsch, G., A. Hortacsu, and X. Lin (2017). Prices and promotions in u.s. retail markets: Descriptive evidence from big data.

Hoch, S. J., E. T. Bradlow, and B. Wansink (1999). The variety of an assortment. *Marketing Science 18*(4), 527–546.

Imai, S., N. Jain, and A. Ching (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica 77*(6), 1865–1899.

IRI (2016). Coffee retail sales share in the united states in 2016, by category.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *The American Economic Review 75*(3), 424–440.

Katz, M. L. and C. Shapiro (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy 94*(4), 822–841.

Kim, J., G. M. Allenby, and P. E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science 21*(3), pp. 229–250.

Kong, P., P. Ellickson, and M. Lovett (2016). *Putting the Horses Before the Cart: Harnessing the Power of Partner Brands.* Ph. D. thesis.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *The American Economic Review 103*(7), 2960–3000.

Li, H. (2016). Intertemporal price discrimination with complementary products: E-books and e-readers.

Melnikov, O. (2013). Demand for differentiated durable products: The case of the u.s. computer printer market. *Economic Inquiry 51*(2), 1277–1298.

Nair, H., P. Chintagunta, and J.-P. Dubé (2004). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics 2*(1), 23–58.

NCA (2016). National coffee drinking trends 2016.

Oi, W. Y. (1971). A disneyland dilemma: Two-part tariffs for a mickey mouse monopoly. *The Quarterly Journal of Economics 85*(1), 77–96.

Rochet, J.-C. and J. Tirole (2003). Platform competition in two-sided markets. *Journal of the European Economic Association 1*(4), 990–1029.

Rochet, J.-C. and J. Tirole (2006). Two-sided markets: A progress report. *The RAND Journal of Economics 37*(3), 645–667.

Rossi, P. E., G. M. Allenby, and R. E. McCulloch (2005). *Bayesian statistics and marketing*. Wiley New York.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999–1033.

Table 8.1: Posterior Estimates of Household Valuation for Coffee

Brand Description	Brand Value	K-Cup Brand Value	Flavored Coffee	Light Roast	Medium Dark	Dark Roast	Assorted Flavors	Kona Blend	Colombian	Sumatra Coffee	Whole Bean
Other Brands	1.693	-	1.143	1.244	1.083	1.128	1.416	1.421	1.333	1.209	1.121
	(1.682, 1.707)	-	(1.122, 1.161)	(1.225, 1.265)	(1.066, 1.102)	(1.105, 1.147)	(1.387, 1.441)	(1.401, 1.446)	(1.311, 1.353)	(1.19, 1.229)	(1.099, 1.142)
CARIBOU KEURIG	-	2.856	1.830	1.992	1.735	1.807	2.268	2.275	2.134	1.937	-
	-	(2.816, 2.918)	(1.798, 1.864)	(1.961, 2.037)	(1.708, 1.772)	(1.774, 1.84)	(2.228, 2.305)	(2.244, 2.334)	(2.102, 2.183)	(1.908, 1.975)	-
CHOCK FULL O NUTS	1.558	-	0.925	1.008	0.878	0.913	1.146	1.152	1.080	0.979	0.907
	(1.547, 1.569)	-	(0.907, 0.945)	(0.991, 1.024)	(0.863, 0.894)	(0.895, 0.934)	(1.119, 1.175)	(1.135, 1.168)	(1.062, 1.099)	(0.962, 0.999)	(0.889, 0.931)
CTL BR	1.727	-	1.203	1.310	1.141	1.188	1.491	1.497	1.404	1.273	1.181
	(1.715, 1.739)	-	(1.178, 1.226)	(1.288, 1.332)	(1.121, 1.16)	(1.164, 1.211)	(1.457, 1.522)	(1.476, 1.519)	(1.38, 1.426)	(1.253, 1.295)	(1.156, 1.206)
DONUT HOUSE KEURIG	-	3.917	2.509	2.731	2.379	2.477	3.106	3.121	2.926	2.654	-
	-	(3.862, 3.975)	(2.467, 2.547)	(2.692, 2.775)	(2.343, 2.415)	(2.433, 2.519)	(3.047, 3.164)	(3.081, 3.173)	(2.881, 2.973)	(2.617, 2.698)	-
DUNKIN' DONUTS	1.667	-	1.095	1.193	1.039	1.081	1.355	1.363	1.278	1.159	1.074
	(1.655, 1.678)	-	(1.074, 1.119)	(1.174, 1.214)	(1.022, 1.058)	(1.06, 1.105)	(1.327, 1.389)	(1.344, 1.383)	(1.257, 1.299)	(1.141, 1.181)	(1.052, 1.1)
EIGHT O'CLOCK	1.582	-	0.962	1.047	0.912	0.949	1.191	1.196	1.122	1.017	0.942
	(1.569, 1.594)	-	(0.94, 0.981)	(1.027, 1.065)	(0.893, 0.928)	(0.927, 0.97)	(1.161, 1.22)	(1.177, 1.215)	(1.1, 1.141)	(0.998, 1.036)	(0.921, 0.966)
FOLGERS	1.839	-	1.431	1.559	1.357	1.413	1.773	1.780	1.670	1.514	1.404
	(1.827, 1.852)	-	(1.405, 1.453)	(1.535, 1.583)	(1.336, 1.378)	(1.388, 1.434)	(1.739, 1.804)	(1.757, 1.81)	(1.645, 1.695)	(1.493, 1.538)	(1.379, 1.429)
FOLGERS KEURIG	-	3.521	2.257	2.455	2.138	2.228	2.796	2.803	2.630	2.388	-
	-	(3.462, 3.605)	(2.211, 2.303)	(2.415, 2.515)	(2.102, 2.187)	(2.183, 2.273)	(2.741, 2.849)	(2.765, 2.885)	(2.586, 2.695)	(2.349, 2.442)	-
GREEN MOUNTAIN KEU	-	3.768	2.414	2.628	2.288	2.384	2.990	3.002	2.815	2.555	-
	-	(3.717, 3.835)	(2.373, 2.451)	(2.591, 2.674)	(2.254, 2.328)	(2.34, 2.42)	(2.937, 3.038)	(2.962, 3.064)	(2.773, 2.865)	(2.519, 2.597)	-
MAXWELL HOUSE	1.756	-	1.258	1.370	1.193	1.243	1.559	1.565	1.468	1.332	1.234
	(1.744, 1.768)	-	(1.235, 1.278)	(1.349, 1.392)	(1.175, 1.212)	(1.219, 1.263)	(1.528, 1.589)	(1.544, 1.59)	(1.445, 1.49)	(1.31, 1.352)	(1.211, 1.258)
NEWMAN'S KEURIG	-	3.058	1.959	2.133	1.857	1.935	2.428	2.437	2.285	2.074	1.923
	-	(3.009, 3.134)	(1.924, 2.002)	(2.097, 2.185)	(1.825, 1.902)	(1.9, 1.974)	(2.383, 2.475)	(2.397, 2.506)	(2.243, 2.343)	(2.036, 2.12)	(1.887, 1.962)
STARBUCKS	1.555	-	0.922	1.004	0.874	0.910	1.142	1.148	1.075	0.976	-
	(1.542, 1.568)	-	(0.901, 0.941)	(0.986, 1.021)	(0.857, 0.89)	(0.889, 0.93)	(1.115, 1.168)	(1.129, 1.165)	(1.054, 1.094)	(0.958, 0.995)	-
STARBUCKS KEURIG	-	3.390	2.173	2.364	2.060	2.146	2.692	2.700	2.533	2.299	2.132
	-	(3.339, 3.462)	(2.131, 2.215)	(2.328, 2.415)	(2.025, 2.103)	(2.102, 2.185)	(2.644, 2.738)	(2.667, 2.766)	(2.493, 2.589)	(2.264, 2.345)	(2.091, 2.171)
TULLY'S KEURIG	-	2.849	1.826	1.988	1.730	1.803	2.264	2.271	2.129	1.933	-
	-	(2.808, 2.918)	(1.795, 1.867)	(1.957, 2.038)	(1.704, 1.774)	(1.772, 1.844)	(2.224, 2.308)	(2.232, 2.337)	(2.095, 2.184)	(1.905, 1.978)	-

Note: 95% posterior coverage region is displayed below the median estimates of preferences. K-Cups come pre-packaged, and thus cannot be of whole bean type.

Table 8.2: Brand Ownership

GMCR Owned Brands	Green Mountain Coffee Roasters, Keurig
Licensed Brands	Caribou Coffee, Newman's Own Organic
Third Party Brands	Folgers, Starbucks, Coffee People, Donut House etc.*

Note: *Many of these brands including Coffee People, Donut House, Tully's Coffee, Timothy's coffee, Diedrich coffee and Van Houtte are later acquired by GMCR. For the purpose of this exercise, I regard them as third party brands as they could very well not be acquired by GMCR if GMCR had adopted a different strategy in supplying K-Cup variety.

Figure 8.1. HHI Before Adoption - At least 5 Purchases

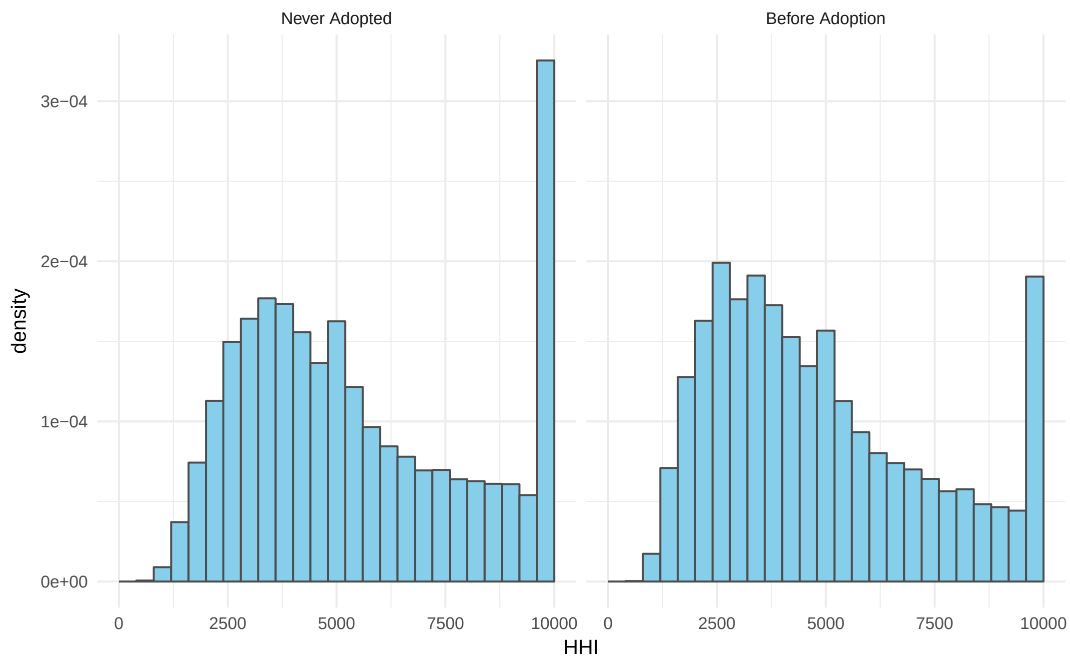


Figure 8.2. Within Person HHI Before/After Adoption

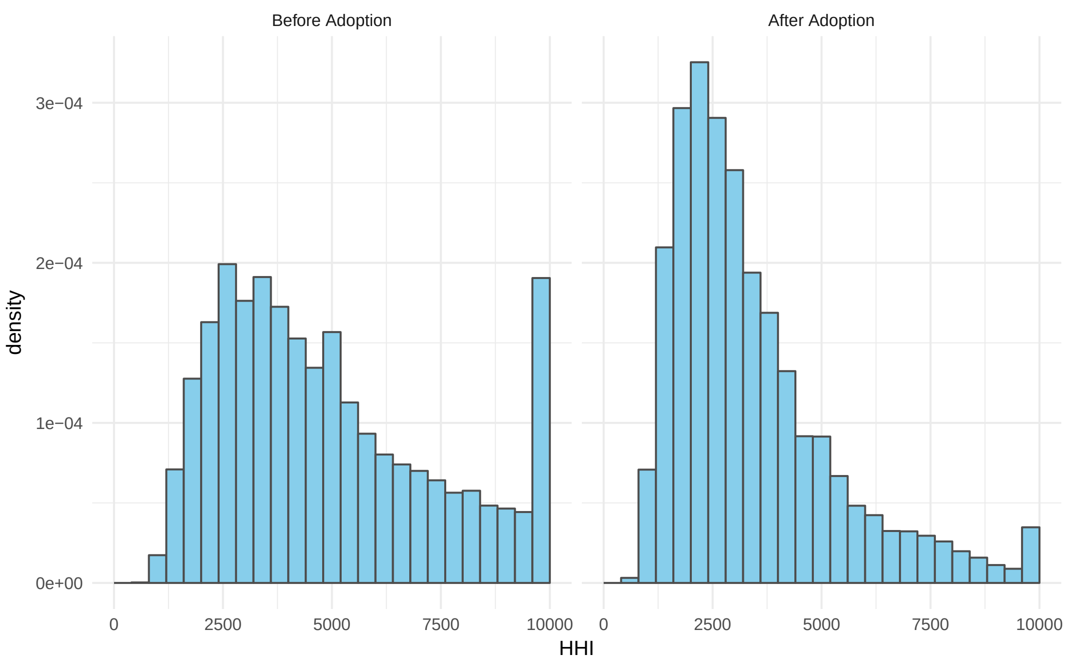


Figure 8.3. HHI Panel - Reference Group

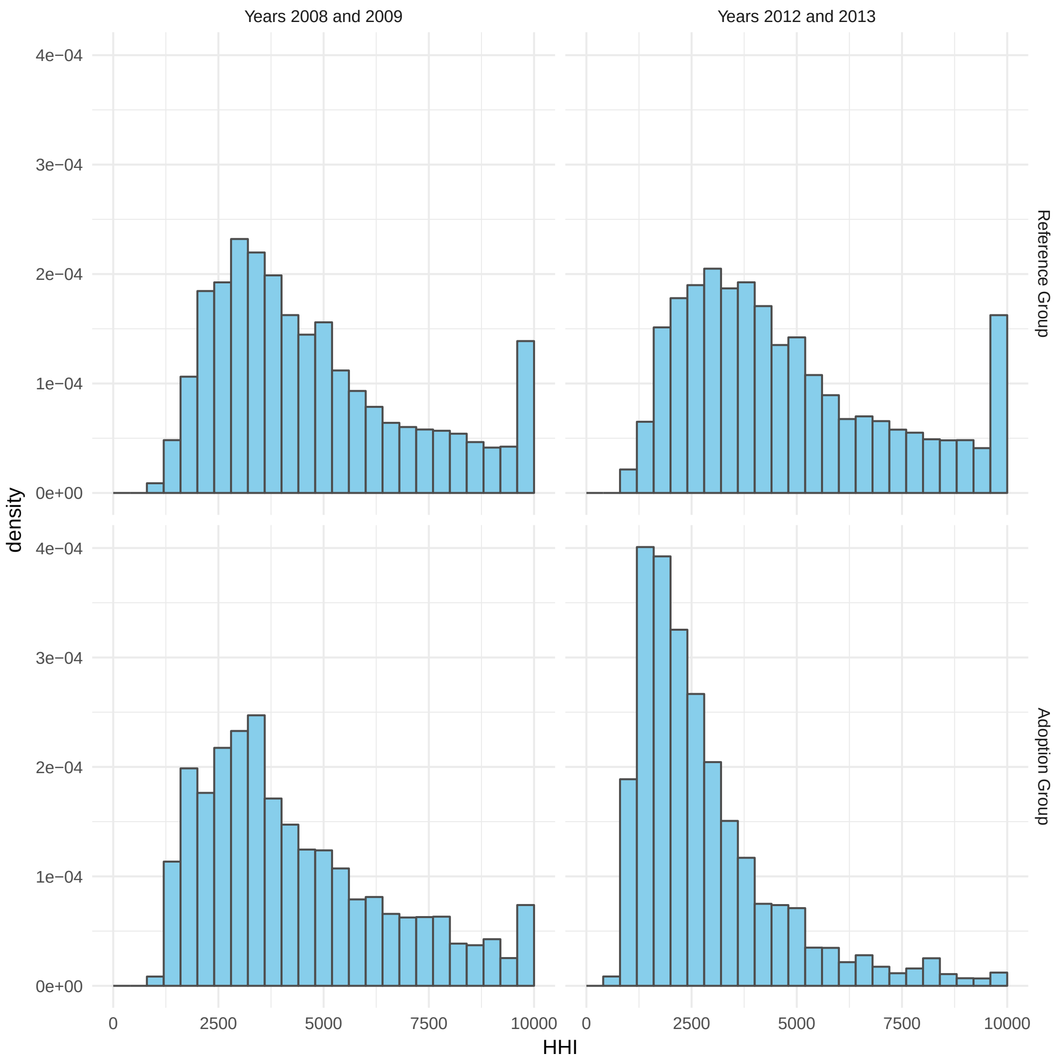


Figure 8.4. Percent of Trips Purchasing 1,2,...,5 brands

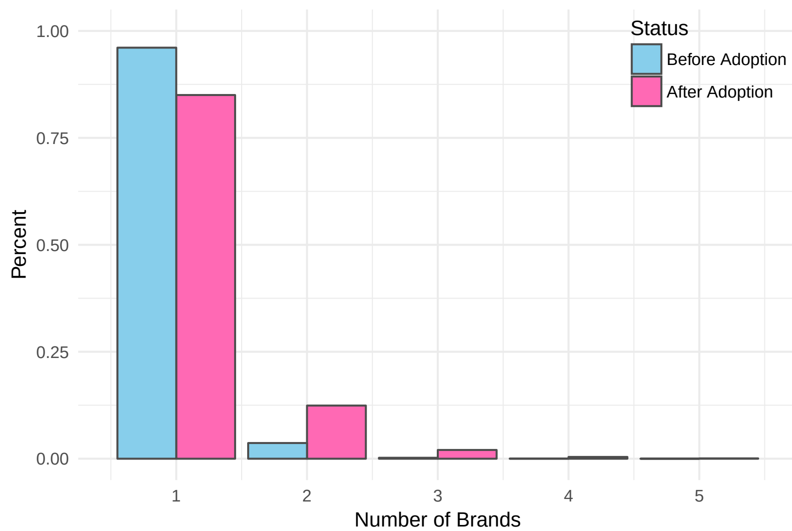


Figure 8.5. Percent of Trips Purchasing a Subset Brands of Last Trip

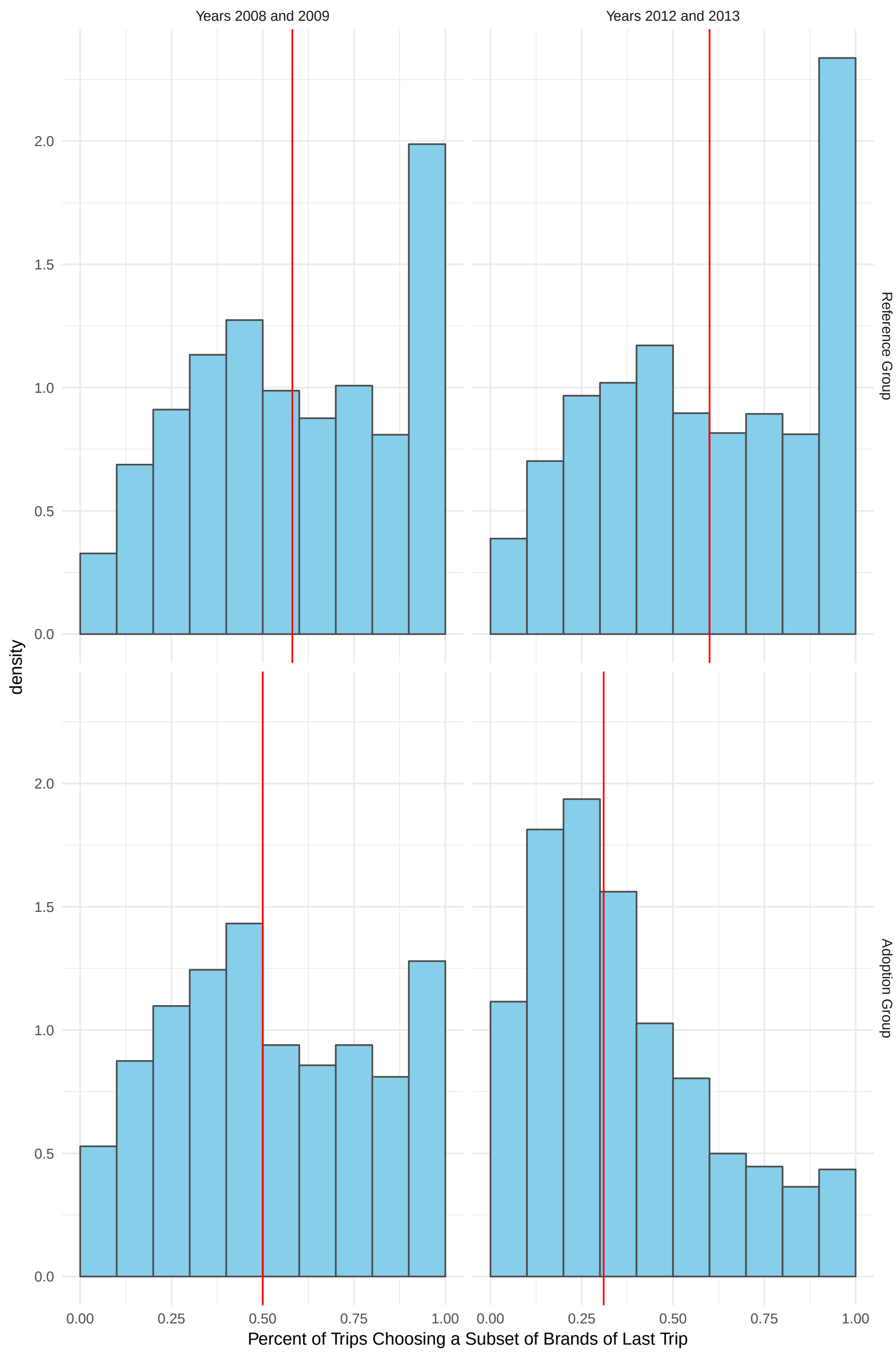


Figure 8.6. HHI Panel (Brand, Flavor and Styles)

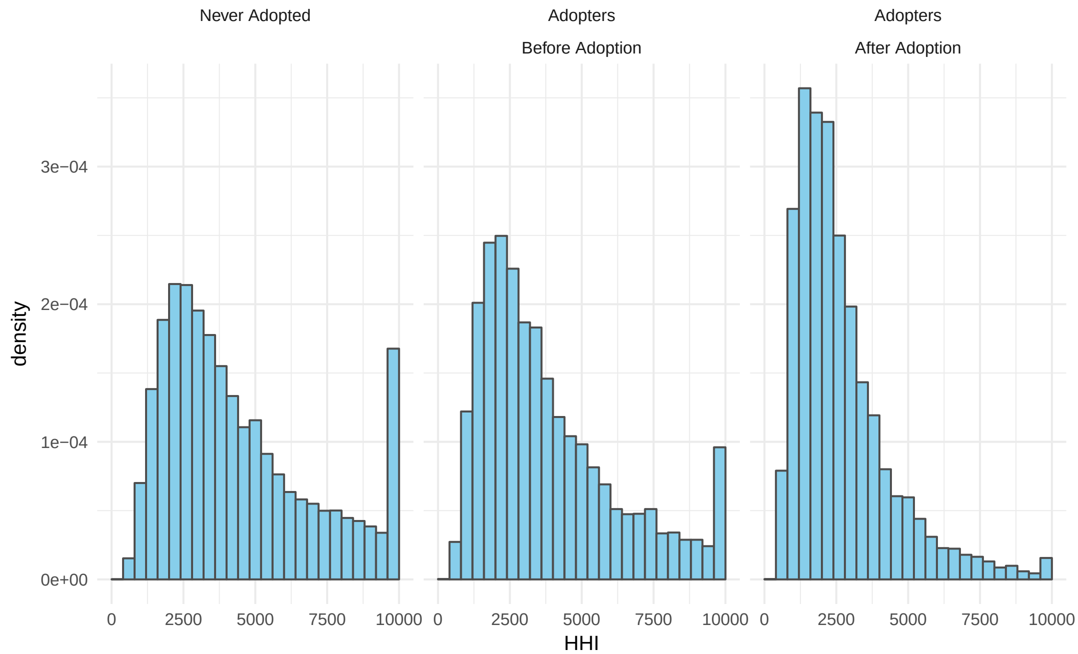


Figure 8.7. Cross-sectional Choice Decomposition

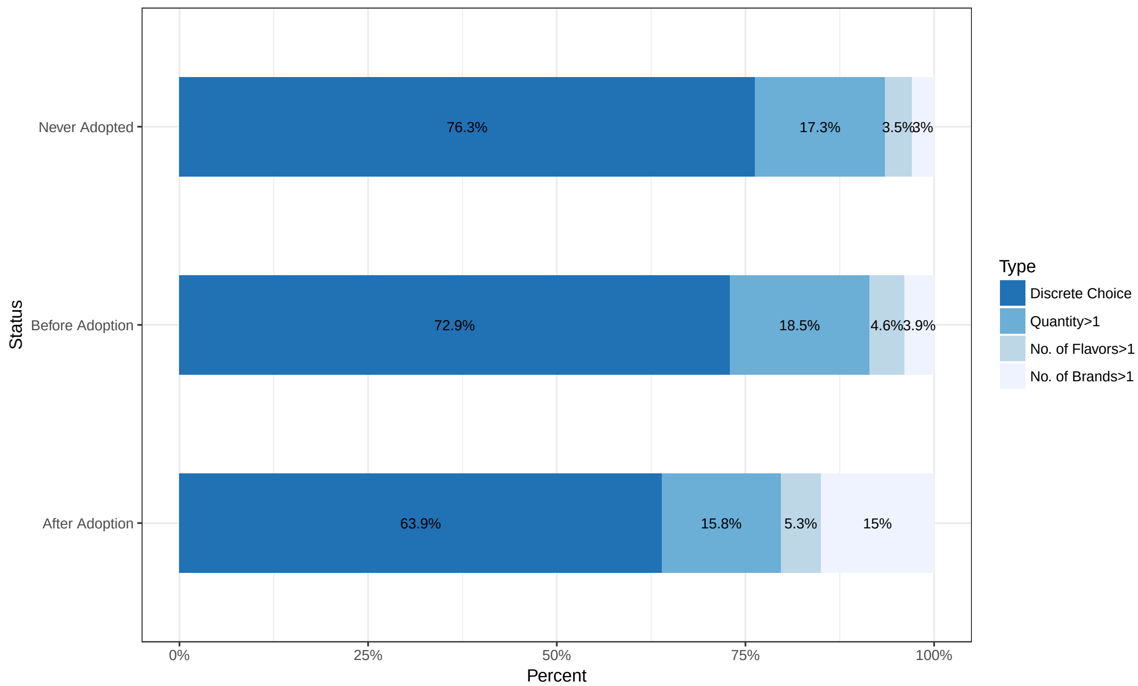


Figure 8.8. Histogram of Trip Portions Purchasing the Same (Subset) Flavor as Last Trip Conditional on No Brand Switching

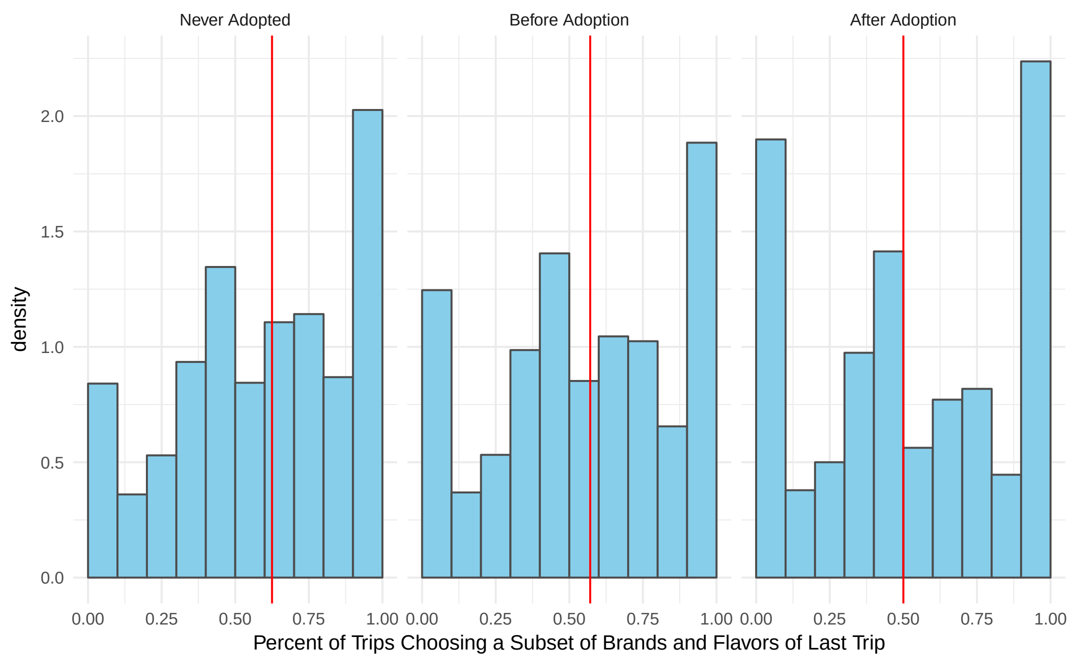


Figure 8.9. Decomposition of Trips, Servings Purchased, and Expenditure by Coffee Type

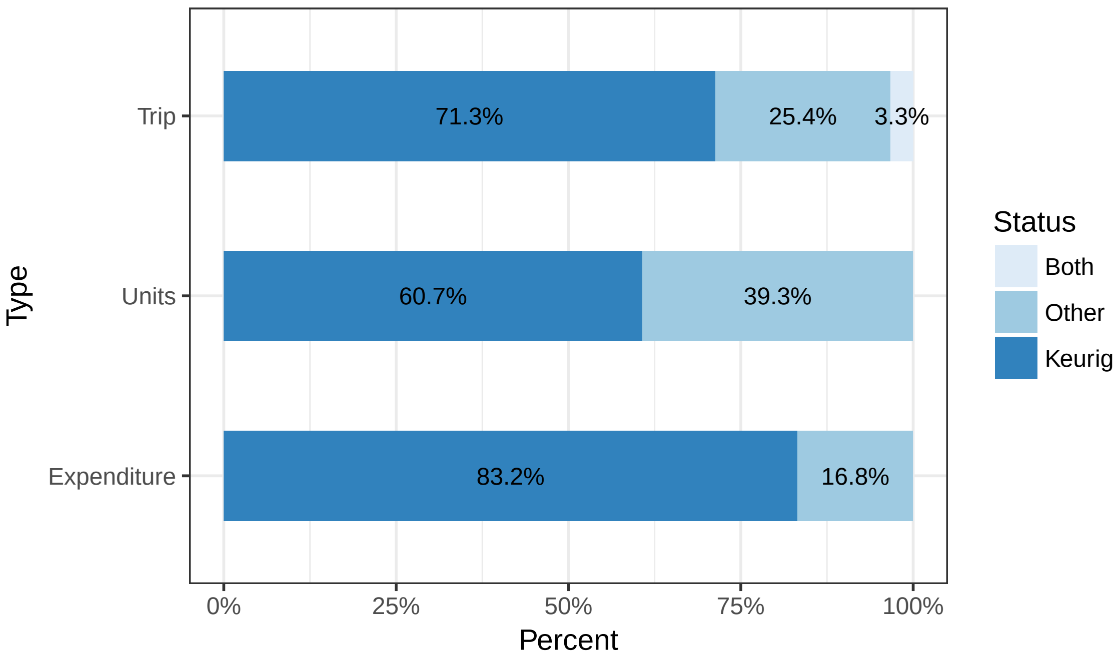


Figure 8.10. Annual Spending on Coffee Including both Ground and K-Cups

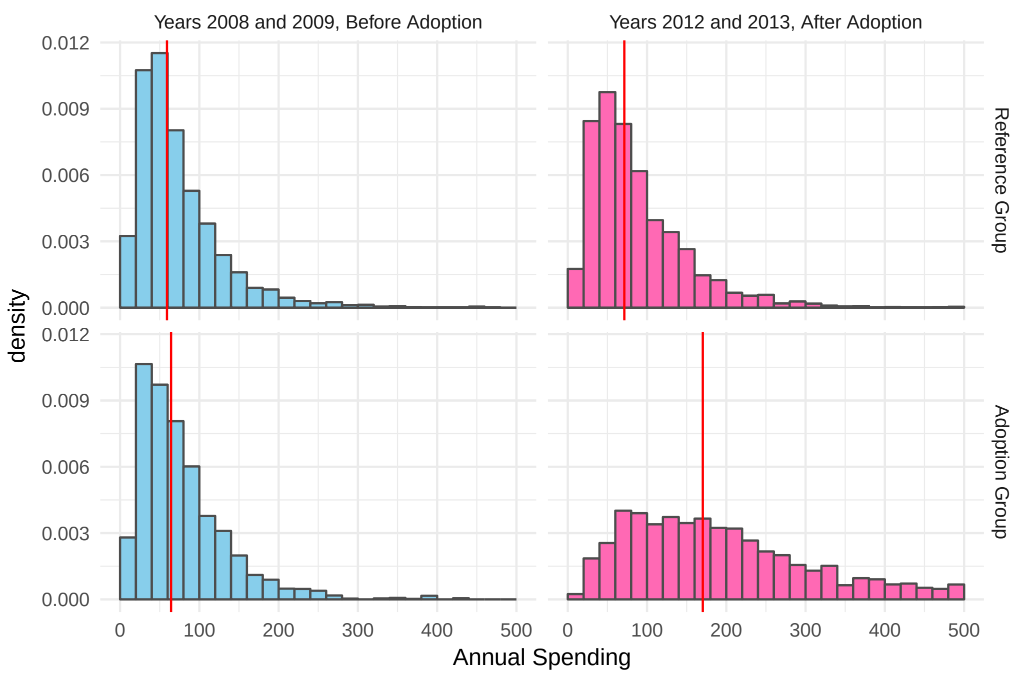


Figure 8.11. Estimated Mean Utility Gain

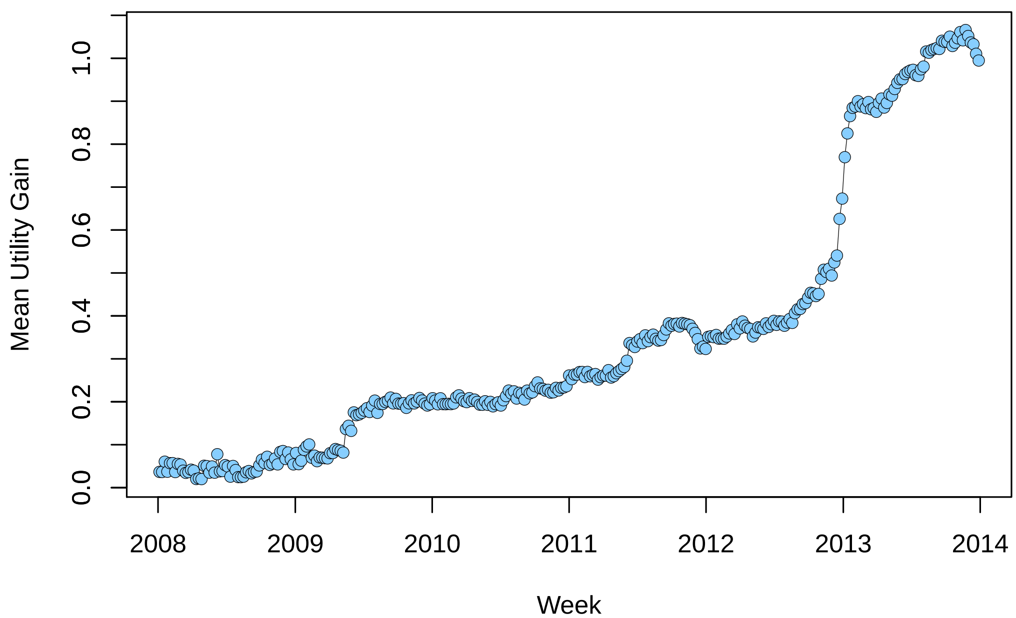


Figure 8.12. Sales Volume and Median Price for Keurig Machine Series in a Large Retailer in New York

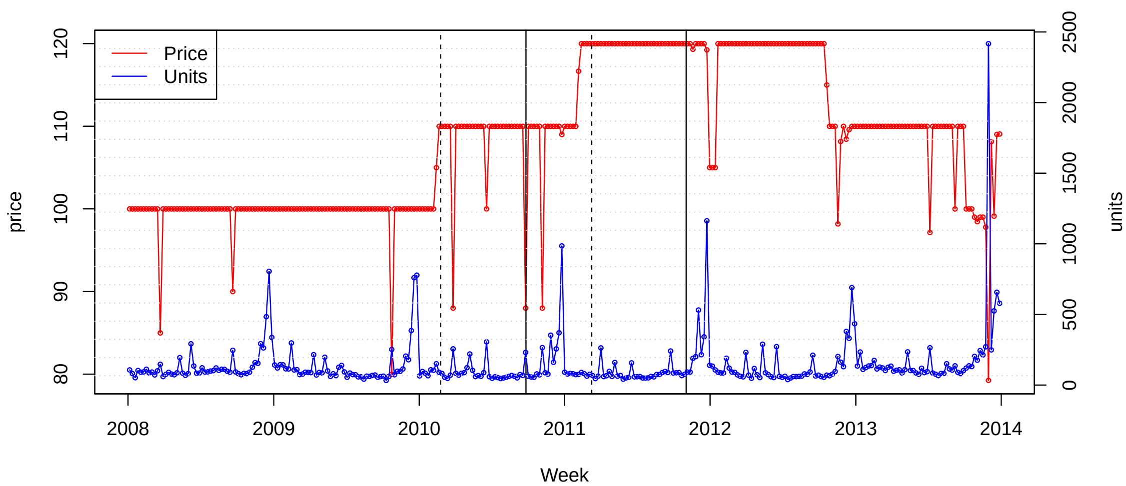
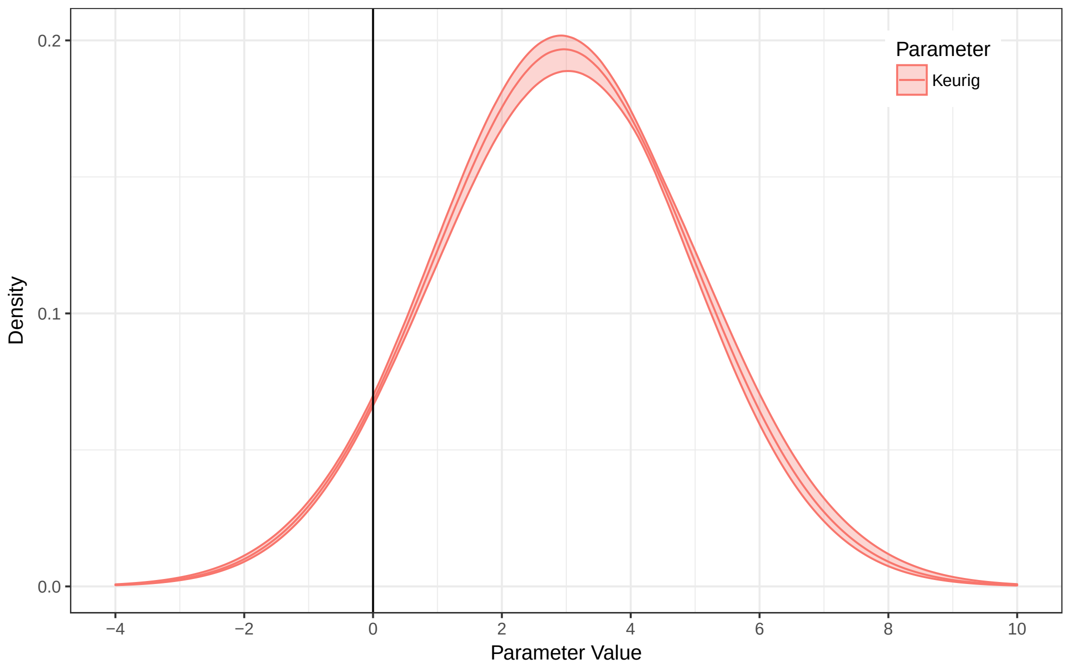
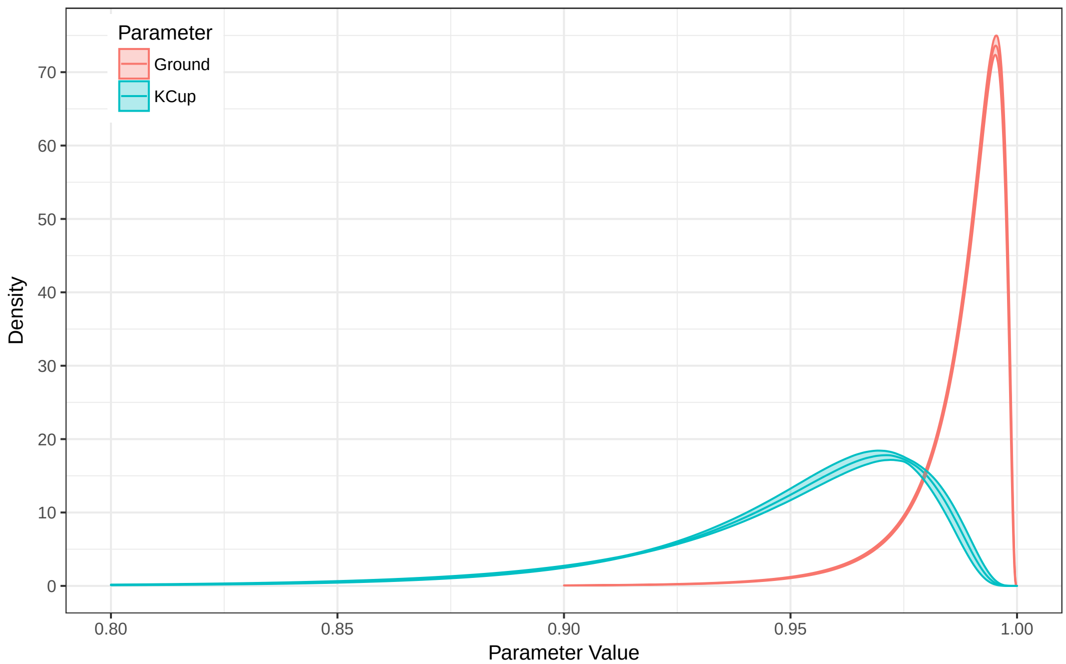


Figure 8.13. K-Cup Fixed Effect, b_1



Note: The band shows 95% posterior credible region.

Figure 8.14. Satiation Parameter α distributions for Ground Coffee and K-Cups



Note: The bands show 95% posterior credible region.

Figure 8.15. State Dependence by Ground Coffee and K-Cup Adjustment

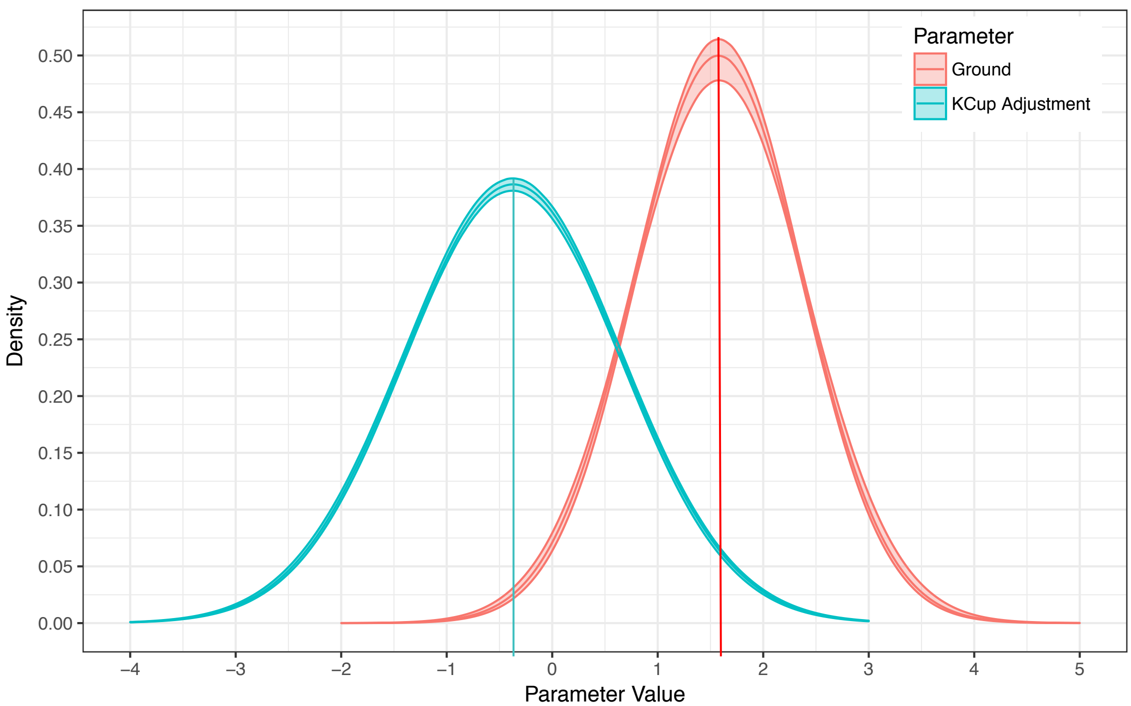
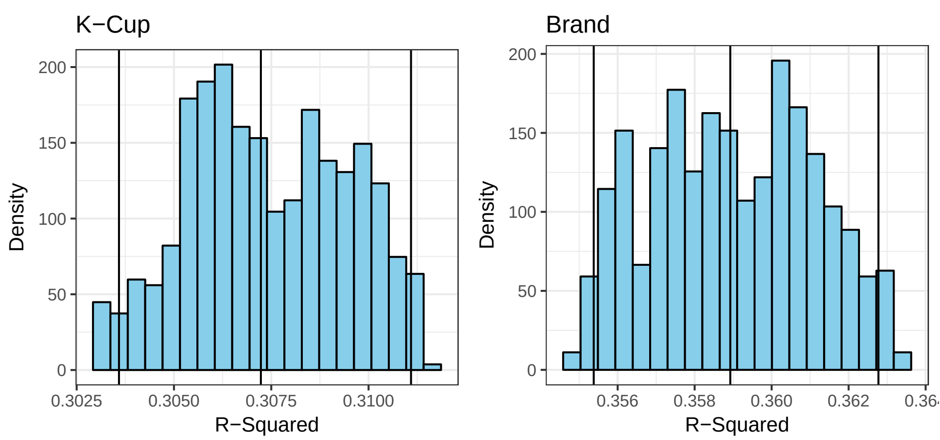
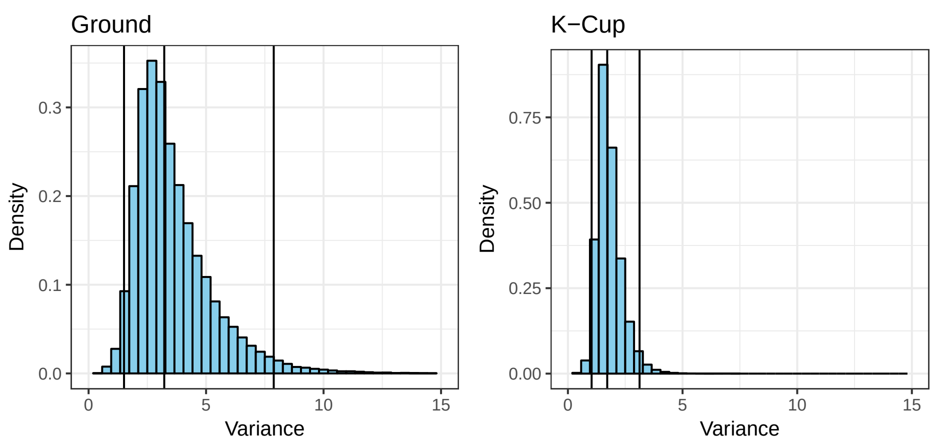


Figure 8.16. R-Squared of Preference Regressions



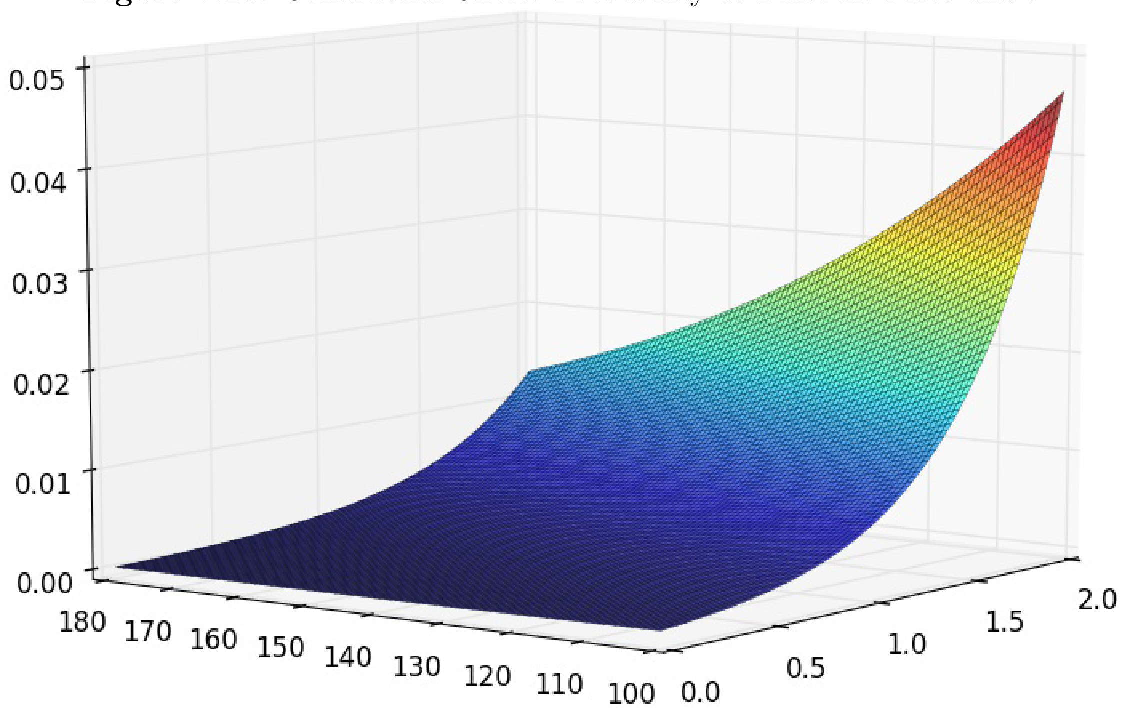
Note: The three vertical lines show 0.025 Percentile, 0.50 percentile, and 0.975 percentile.

Figure 8.17. Within Household Variance Analysis



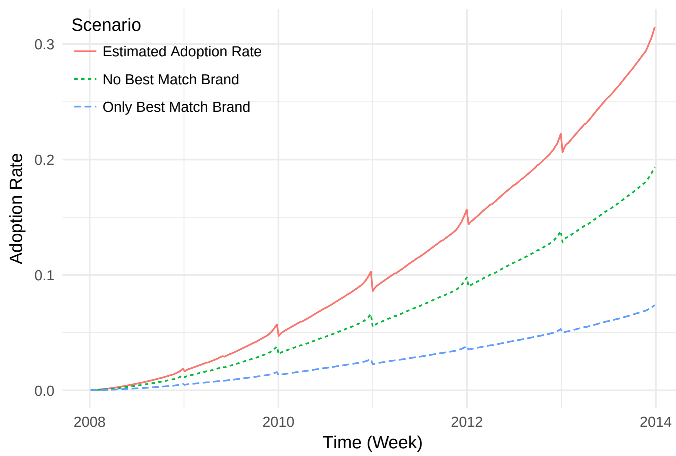
Note: The three vertical lines show 0.025 Percentile, 0.50 percentile, and 0.975 percentile.

Figure 8.18. Conditional Choice Probability at Different Price and δ



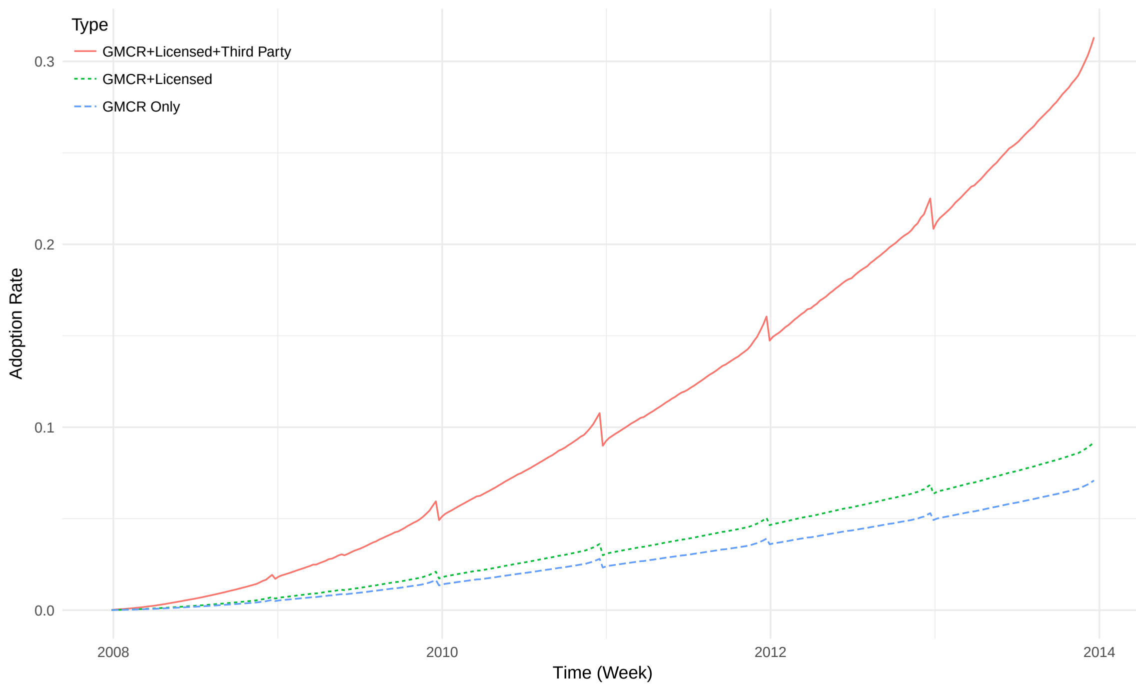
Note: The vertical axis the choice probability, the axis from 180 to 100 is the price, the the axis from 0 to 2.0 is the utility gain δ, and the mean price is constrained at \$140.

Figure 8.19. Expected Keurig Adoption Rate over Time by Types of K-Cups Supplied



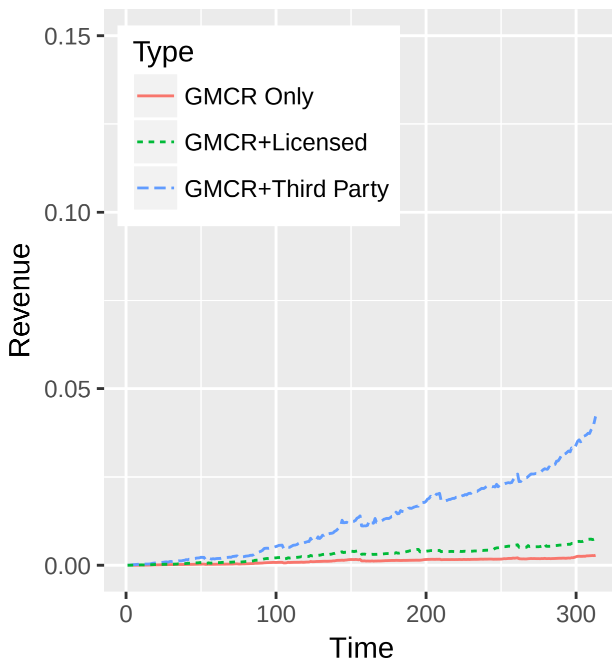
Note: The sharp increase/decrease are due to seasonality and burnin period for households newly entering the consumer panel.

Figure 8.20. Expected Adoption Rate by Variety Supplied



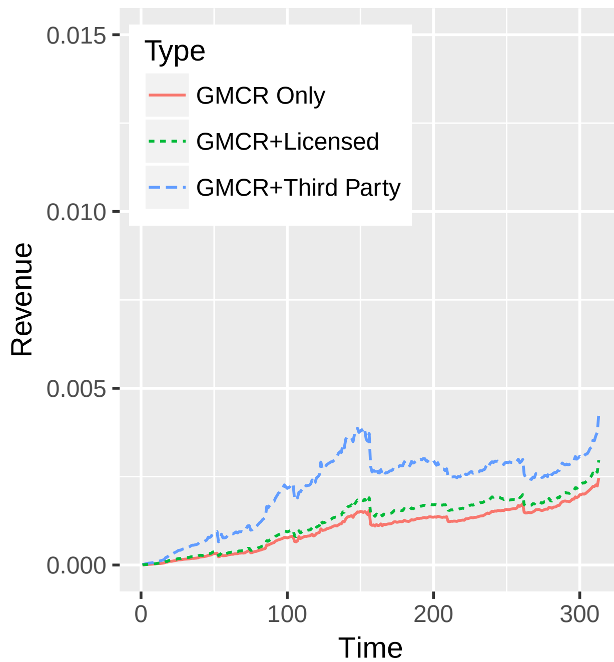
Note: The sharp increase/decrease are due to seasonality and burnin period for households newly entering the consumer panel.

Figure 8.21. Overall Expected Revenue by Variety Supplied



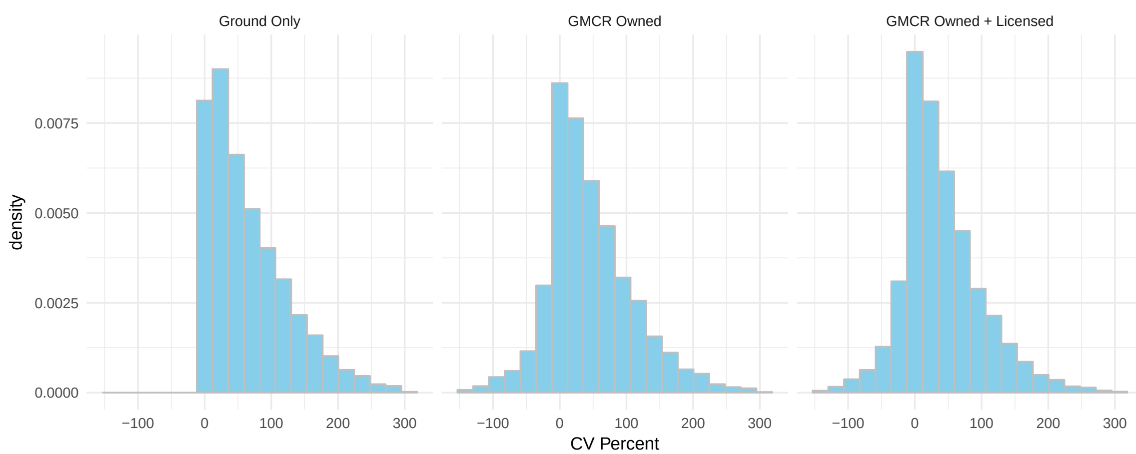
Note: The sharp increase/decrease are due to seasonality and burnin period for households newly entering the consumer panel.

Figure 8.22. GMCR Owned Brand Expected Revenue by Variety Supplied



Note: The sharp increase/decrease are due to seasonality and burnin period for households newly entering the consumer panel.

Figure 8.23. Compensating Value Percentage by Variety Supplied



Note: Compensating values are computed relative to the true evolution of variety on the platform.

APPENDIX A

LIKELIHOOD DERIVATION

The direct utility function in terms of expenditures,

$$U(e_0, e_1, \ldots, e_K) = \psi_0 e_0 + \left\{ \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left[\left(\frac{e_k}{p_k \gamma_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho},$$

subject to budget constraint

$$\sum_{k=0}^{K} e_k = E.$$

Then the Lagrangian,

$$\mathcal{L}(e_0, e_1, \ldots, e_K) = \psi_0 e_0 + \left\{ \underbrace{\sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left[\left(\frac{e_k}{p_k \gamma_k} + 1 \right)^{\alpha_k} - 1 \right]}_{\Xi} \right\}^{\rho} + \lambda \left(E - \sum_{k=0}^{K} e_k \right)$$

Assume the budget is big enough, and the outside option is always chosen. Therefore, for optimal allocation, the optimal expenditures $e_1^*, \ldots, e_K^*$ must meet the Kuhn-Tucker condition, and the conditions are

$$\begin{aligned} \psi_0 - \lambda &= 0 \\ \rho \Xi^{\rho-1} \frac{\psi_k}{p_k} \left(\frac{e_k^*}{p_k \gamma_k} + 1 \right)^{\alpha_k - 1} - \lambda &= 0 \text{ if } e_k^* > 0 \\ \rho \Xi^{\rho-1} \frac{\psi_j}{p_j} \left(\frac{e_j^*}{p_j \gamma_j} + 1 \right)^{\alpha_j - 1} - \lambda &\leq 0 \text{ if } e_j^* = 0. \end{aligned}$$

In this set up, $\lambda > 0$ and I can take logs on both side of the conditions,

$$\begin{aligned}
\ln\lambda &= \rho\varepsilon_0 \\
\ln\rho + (\rho-1)\ln\Xi + z_k\beta + \varepsilon_k + (\alpha_k - 1)\ln\left(\frac{e_k^*}{p_k\gamma_k}+1\right) - \ln p_k &= \ln\lambda \text{ if } e_k^* > 0 \\
\ln\rho + (\rho-1)\ln\Xi + z_j\beta + \varepsilon_j + \left(\alpha_j - 1\right)\ln\left(\frac{e_j^*}{p_j\gamma_j}+1\right) - \ln p_j &\leq \ln\lambda \text{ if } e_j^* = 0.
\end{aligned}$$

Re-order the products so that the first $1,\ldots,M$ products are the chosen ones,

$$\begin{aligned}
(\rho-1)\ln\Xi + \ln\rho + V_1 + \varepsilon_1 &= \rho\varepsilon_0 \\
V_1 + \varepsilon_1 &= V_k + \varepsilon_k \text{ if } k = 2,\ldots,M \\
V_1 + \varepsilon_1 &\geq V_k + \varepsilon_k \text{ if } k = M+1,\ldots,K.
\end{aligned}$$

where $V_k = z_k\beta + (\alpha_k - 1)\ln\left(\frac{e_k^*}{p_k\gamma_k}+1\right) - \ln p_k$. Now, let's simplify Ξ,

$$\begin{aligned}
\Xi &= \sum_{k=1}^{M}\frac{\gamma_k}{\alpha_k}\psi_k\left[\left(\frac{e_k^*}{p_k\gamma_k}+1\right)^{\alpha_k} - 1\right] + \sum_{k=M+1}^{K}\frac{\gamma_k}{\alpha_k}\psi_k\left[(0+1)^{\alpha_k} - 1\right] \\
&= \sum_{k=1}^{M}\frac{\gamma_k}{\alpha_k}\psi_k\left[\left(\frac{e_k^*}{p_k\gamma_k}+1\right)^{\alpha_k} - 1\right] \\
&= \sum_{k=1}^{M}\frac{\gamma_k}{\alpha_k}\exp\left(z_k\beta + \varepsilon_k\right)\left[\left(\frac{e_k^*}{p_k\gamma_k}+1\right)^{\alpha_k} - 1\right] \\
&= \sum_{k=1}^{M}\frac{\gamma_k}{\alpha_k}\exp\left(V_1 - V_k + z_k\beta + \varepsilon_1\right)\left[\left(\frac{e_k^*}{p_k\gamma_k}+1\right)^{\alpha_k} - 1\right] \\
&= \exp\left(V_1 + \varepsilon_1\right)\underbrace{\sum_{k=1}^{M}\frac{\gamma_k}{\alpha_k}\exp\left(z_k\beta - V_k\right)\left[\left(\frac{e_k^*}{p_k\gamma_k}+1\right)^{\alpha_k} - 1\right]}_{\Omega}.
\end{aligned}$$

In log,

$$\ln \Xi = V_1 + \varepsilon_1 + \ln \Omega.$$

Given Ω and Ξ, the first order condition simplifies to,

$$(\rho - 1)(V_1 + \varepsilon_1 + \ln \Omega) + \ln \rho + V_1 + \varepsilon_1 = \rho \varepsilon_0,$$

$$V_1 + \varepsilon_1 + \frac{(\rho - 1)}{\rho} \ln \Omega + \frac{\ln \rho}{\rho} = \varepsilon_0.$$

Reordering the variables implies,

$$\begin{aligned}
\varepsilon_k &= \frac{(1-\rho)}{\rho} \ln \Omega - \frac{\ln \rho}{\rho} - V_k + \varepsilon_0 \text{ if } k = 1, \ldots, M \\
\varepsilon_k &\leq \frac{(1-\rho)}{\rho} \ln \Omega - \frac{\ln \rho}{\rho} - V_k + \varepsilon_0 \text{ if } k = M+1, \ldots, K.
\end{aligned}$$

Given the above equations, inequalities and the independent error assumption, the probability of the expenditures is written as follows,

$$
\begin{aligned}
\Pr(e^*) &= |J| \int_{\varepsilon_0=-\infty}^{\infty} \int_{\varepsilon_{M+1}=-\infty}^{\frac{(1-\rho)}{\rho}+\frac{\ln\rho}{\rho}\ln\Omega - V_{M+1}+\varepsilon_0} \cdots \int_{\varepsilon_K=-\infty}^{\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}-V_K+\varepsilon_0} \\
&\quad f(\varepsilon_0) \prod_{k=1}^{M} f\left(\frac{(1-\rho)}{\rho}\ln\Omega - \frac{\ln\rho}{\rho} - V_k + \varepsilon_0\right) \\
&\quad \times f(\varepsilon_{M+1}) \cdot \cdots \cdot f(\varepsilon_K)\, d\varepsilon_K \ldots d\varepsilon_{M+1}. \\
&= |J| \int_{\varepsilon_0=-\infty}^{\infty} f(\varepsilon_0) \prod_{k=1}^{M} e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_k-\varepsilon_0} e^{-e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_k-\varepsilon_0}} \cdot \\
&\quad e^{-e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_{M+1}-\varepsilon_0}} \ldots e^{-e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_K-\varepsilon_0}} d\varepsilon_0 \\
&= |J| \left[\prod_{i=1}^{M} e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_i}\right] \cdot \\
&\quad \left[\int_{\varepsilon_0=-\infty}^{\infty} \left(e^{-\varepsilon_0}\right)^M e^{-e^{-\varepsilon_0}\left[1+\sum_{j=1}^{K} e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_j}\right]} e^{-\varepsilon_0} d\varepsilon_0\right] \\
&= |J| \left[\prod_{i=1}^{M} e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_i}\right] \left\{\frac{1}{\left[1+\sum_{j=1}^{K} e^{-\frac{(1-\rho)}{\rho}\ln\Omega+\frac{\ln\rho}{\rho}+V_j}\right]^{M+1}}\right\} M! \\
&= |J| \left\{\frac{e^{\frac{(1-\rho)}{\rho}\ln\Omega-\frac{\ln\rho}{\rho}} \prod_{i=1}^{M} e^{V_i}}{\left[e^{\frac{(1-\rho)}{\rho}\ln\Omega-\frac{\ln\rho}{\rho}} + \sum_{j=1}^{K} e^{V_j}\right]^{M+1}}\right\} M!,
\end{aligned}
$$

where J is the Jacobian for the change of variable.

A.1 Jacobian Derivation

The i, hth element of J is

$$J_{ih} = \frac{\partial\left[\frac{(1-\rho)}{\rho}\ln\Omega - \frac{\ln\rho}{\rho} - V_i + \varepsilon_0\right]}{\partial e_h} = \frac{\partial\left[\frac{(1-\rho)}{\rho}\ln\Omega - V_i\right]}{\partial e_h}.$$

Let

$$\begin{aligned}
\frac{\partial\Omega}{\partial e_i} = c_i &= \frac{\gamma_i}{\alpha_i}\Bigg\{-\exp\left(Z_i\beta - V_i\right)\frac{\alpha_i - 1}{e_i^* + p_i\gamma_i}\left[\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i} - 1\right] + \\
&\quad \frac{\alpha_i}{p_i\gamma_i}\exp\left(Z_i\beta - V_i\right)\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i - 1}\Bigg\} \\
&= \frac{\gamma_i}{\alpha_i}\exp\left(Z_i\beta - V_i\right)\left\{\frac{\alpha_i}{p_i\gamma_i}\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i-1} - \frac{\alpha_i - 1}{e_i^* + p_i\gamma_i}\left[\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i} - 1\right]\right\} \\
&= \frac{\gamma_i}{\alpha_i}\exp\left(Z_i\beta - V_i\right)\left\{\frac{\alpha_i}{p_i\gamma_i}\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i-1} - \frac{\alpha_i - 1}{e_i^* + p_i\gamma_i}\left[\left(\frac{e_i}{p_i\gamma_i}+1\right)^{\alpha_i} - 1\right]\right\}.
\end{aligned}$$

Given the derivation,

$$J_{ii} = \frac{(1-\rho)c_i}{\rho\Omega} - \frac{\alpha_i - 1}{e_i + p_i\gamma_i}$$

and

$$J_{ij} = \frac{(1-\rho)c_j}{\rho\Omega} \quad i \neq j.$$

Therefore,

$$|J| = \begin{vmatrix}
\frac{(1-\rho)c_1}{\rho\Omega} - \frac{\alpha_1-1}{e_1+p_1\gamma_1} & \frac{(1-\rho)c_2}{\rho\Omega} & \cdots & \frac{(1-\rho)c_M}{\rho\Omega} \\
\frac{(1-\rho)c_1}{\rho\Omega} & \frac{(1-\rho)c_2}{\rho\Omega} - \frac{\alpha_2-1}{e_2+p_2\gamma_2} & \cdots & \frac{(1-\rho)c_M}{\rho\Omega} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{(1-\rho)c_1}{\rho\Omega} & \frac{(1-\rho)c_2}{\rho\Omega} & \cdots & \frac{(1-\rho)c_M}{\rho\Omega} - \frac{\alpha_M-1}{e_M+p_M\gamma_M}
\end{vmatrix}.$$

APPENDIX B

BAYESIAN ALGORITHMS FOR ESTIMATION

B.1 Coffee Purchase Estimation

Following Rossi et al. (2005), define

$$\begin{aligned}
\hat{\Delta} &= (Z'Z)^{-1}Z'\Theta \quad \Theta = [\theta_{1c}, \theta_{2c}, \ldots, \theta_{nc}]' \\
\tilde{\Delta} &= (Z'Z + A)^{-1}(Z'Z\hat{\Delta} + A\bar{\Delta}) \\
S &= (\Theta - Z\tilde{\Delta})'(\Theta - Z\tilde{\Delta}) + (\tilde{\Delta} - \bar{\Delta})'A(\tilde{\Delta} - \bar{\Delta})
\end{aligned}$$

and given the conjugacy, the posterior distributions of Σ and $\mathrm{vec}(\Delta)$ are

$$\begin{aligned}
\Sigma|\Theta, Z &\sim IW(\nu_0 + n, V_0 + S) \\
\mathrm{vec}(\Delta)|\Theta, Z, \Sigma &\sim MVN(\mathrm{vec}(\tilde{\Delta}), \Sigma \otimes (Z'Z + A)^{-1}).
\end{aligned}$$

After these draws, the conditional prior of Θ is

$$\mathrm{vec}(\Theta)|\Delta, Z, \Sigma \sim MVN(\mathrm{vec}(Z\Delta), \Sigma_{\boldsymbol{\beta}} \otimes I_n),$$

and for each individual θ_{hc},

$$p(\theta_{hc}|\Delta, Z_i, \Sigma) = MVN(\Delta' Z_h, \Sigma).$$

Then, I can draw individual θ_{hc} from its posterior using MH method,

$$p(\theta_{hc}|D_h, Z_h, \Delta, \Sigma) \propto CL_h(\theta_{hc}|D_h) \cdot p(\theta_{hc}|Z_h, \Delta, \Sigma).$$

Then the algorithm can then cycle through these steps to construct the MCMC draws.

B.2 Machine Adoption Estimation

The algorithm in my context involves the following steps:

- Initialize a set of parameters and state pairs in their feasible ranges.
  - Approximate the expected value function at each parameter and state pair using a non-parametric kernel method over parameters and non-varying states. For varying states, use the density transition for states as the kernel. This essentially gives the expected value function $\mathbb{E}\left[\mathcal{W}(0, P', \bar{P}', \delta')\right]$.
  - Use Bellman iteration to update $\mathcal{W}(0, P, \bar{P}, \delta)$.
  - Repeat the process 100 times – only crude estimates are needed.
- Use a random walk MH algorithm to draw the parameters.
  - Draw θ_a^l, and a uniform draw over the range of states s^l, and then use the Bellman equation to update $\mathcal{W}(0, s^l|\theta_a^l)$.
  - Compute the likelihood and use MH to update θ_a.

In this way, the algorithm successively has better estimates of the expected value function while having better estimates of parameters.

APPENDIX C

ALGORITHM FOR COMPUTING INDIRECT UTILITY

Given the set up of the model, the marginal utility of an extra dollar spent on product k is

$$\rho \Xi^{\rho-1} \frac{\psi_k}{p_k} \left(\frac{e_k^*}{p_k \gamma_k} + 1 \right)^{\alpha_k - 1}.$$

where Ξ is defined the same as in section A. First I show the marginal utility decreases with the expenditure. Take derivative of the marginal utility with respect to the expenditure on product k, which gives

$$\rho(\rho-1)\Xi^{\rho-2} \frac{\alpha_k \psi_k}{p_k^2 \gamma_k} \left(\frac{e_k}{p_k \gamma_k} + 1 \right)^{2\alpha_k - 2} + (\alpha_k - 1)\rho \Xi^{\rho-1} \frac{\psi_k}{p_k^2 \gamma_k} \left(\frac{e_k^*}{p_k \gamma_k} + 1 \right)^{\alpha_k - 2},$$

which is less than 0 since $\rho < 1$ and $\alpha_k < 1$. Therefore, the marginal utility of the product decreases with increasing expenditure. In addition, notice at least one inside good will be chosen given the structure of the utility. Given the first order conditions of the utility maximization problem and smoothness of the marginal utility, the product with highest marginal utility when all the expenditures goes to 0, which implies

$$z_1 \beta - \ln p_1 + \varepsilon_1 \geq z_k \beta - \ln p_k + \varepsilon_k, \; k = 1, \ldots, K,$$

is always chosen. In addition, a second inside product will only be chosen if when the expenditure on product 1 meets the following condition,

$$z_1 \beta + (\alpha_1 - 1) \ln \left(\frac{e_1^*}{p_1 \gamma_1} + 1 \right) - \ln p_1 + \varepsilon_1 = z_2 \beta - \ln p_2 + \varepsilon_2,$$

where the index 2 denotes the second highest marginal utility when all expenditures goes to 0. Given such structure, we can solve the problem using Algorithm C.1[1]. The algorithm only several simple calculations and one root calculation to solve a large constrained optimization problem. With 100 products, I achieve a speed up of 30-50 times, which enables me to use Monte Carlo method to simulate the indirect utility and thus utility gain.

1. Note, given the optimality condition,

$$e_2^{\#}(e_1^{\#}) = p_2\gamma_2 \left[\exp\left(\frac{u_1(0) + (\alpha_1 - 1)\ln\left(\frac{e_1^{\#}}{p_1\gamma_1} + 1\right) - u_2(0)}{\alpha_2 - 1} \right) - 1 \right].$$

Then, the problem can be solved via solving,

$$\ln\rho + (\rho - 1)\ln\Xi\left(e_1^{\#}, e_2^{\#}(e_1^{\#}), \ldots, 0\right) + u_1(e_1^{\#}) = \rho\varepsilon_0.$$

Algorithm C.1 Computing the Maximum

1. Pre-compute
$$\delta_k = z_k\beta - \ln p_k$$
for each $k = 1, \ldots, K$.

2. Draw ε_k, $k = 0, \ldots, K$, and compute the "quasi" marginal utility at $e_k^* = 0$ for each choice k,
$$u_k(0) = \delta_k + \varepsilon_k + (\alpha_k - 1)\ln\left(\frac{0}{p_k\gamma_k} + 1\right) = \delta_k + \varepsilon_k.$$
Order the choices based on $u_k(0)$, and $u_1(0) \geq u_2(0) \geq \cdots \geq u_K(0)$. Also, the equation implies $\psi_k = \exp\left[u_k(0) + \ln p_k\right]$.

3. Find e_1^* such that
$$u_1(e_1^*) = u_2(0),$$
which is given by
$$e_1^* = p_1\gamma_1\left[\exp\left(\frac{u_2(0) - u_1(0)}{\alpha_1 - 1}\right) - 1\right].$$

4. Given e_1^*, compute
$$\zeta = \ln\rho + (\rho - 1)\ln\Xi\left(e_1^*, 0, \ldots, 0\right) + u_1(e_1^*).$$

5. If $\zeta \leq \rho\varepsilon_0$, stop the algorithm and solve for $e_1^{\#}$ by solving,
$$\ln\rho + (\rho - 1)\ln\Xi\left(e_1^{\#}, 0, \ldots, 0\right) + u_1(e_1^{\#}) = \rho\varepsilon_0$$

6. If $\zeta > \rho\varepsilon_0$, find e_1^* and e_2^* (or add another product) such that
$$\begin{aligned} u_1(e_1^*) &= u_3(0) \\ u_1(e_2^*) &= u_3(0) \end{aligned}$$

7. which is given by
$$e_j^* = p_j\gamma_j\left[\exp\left(\frac{u_3(0) - u_j(0)}{\alpha_j - 1}\right) - 1\right] \quad j = 1, 2.$$

8. Given the the solved expenditures, re-compute ζ and repeat 5, 6 and 7.

APPENDIX D

HOUSEHOLD IDENTIFICATION

For the purpose of identifying potential Keurig adopters, I constrain to households whom have purchased in the ground coffee category in the Nielsen data. Keurig adopters without prior purchases of ground coffee are rare. I define an adopter as one who made at least two purchases of compatible portion packs, and who purchased in a span of at least 30 days. For Keurig-machine owners, the Homescan panel has 12,950 adoption households if I constrain to the sample to two K-Cups, and 11,782 if three K-Cups. Because RMS price data is only available from 2006 and onward, I drop all households who are not present after 2006.

In addition to K-Cup purchases, I also consider purchase rate - average number of purchases and quantity purchased for K-cups as well as other products before and after adoption. I compute the consumption rate before and after based on the number of purchases. I only include households who make at least 1.5 purchases annually or 1.25 purchases after adoption. The criteria is a rather relaxed since most do make more purchases, and the filtration eliminates very casual coffee households which are about 22% of the Homescan panel of households who ever purchased any coffee.

APPENDIX E

GROUND COFFEE AND K-CUP PRICES IMPUTATION

Keurig machine and K-Cups are distributed through many different channels including grocery, mass merchandiser, department stores, home furnishing stores, online retailers and even electronic stores. The retail management system is a comprehensive retailer panel, and mainly cover grocery stores, drug stores, convenience stores, dollar stores, and mass merchandisers. The retail management system is assumed to be a more accurate description of the retail environment since it records of the comprehensive sales and average prices at store week level. So, all sales through a store are captured in the data. On the other hand, the Homescan data is a national panel of 60,000+ households, but they are sparsely populated considering the size of the United States. Capturing the retail environment for all market chains are nearly impossible in the Homescan data. At such, I only attempt to capture retail environment at top retailers of coffee in Homescan data and combine it with RMS data to create a comprehensive retail environment data set. Top retailers in HMS is defined based on Keurig machine adopters. We constrain it to top 30 markets, which cover about 58% of the total adoption households and 51% of the total coffee consumption households.

Before 2008, Keurig machine holding in the Homescan dataset is minimal, I filter the data and only build the data set from 2008 onward. As data is sparse (even just 1 or 2 coffee consuming households in Homescan) for the smaller markets, I constrain my analysis to top 30 markets by the number of coffee consuming households. These DMAs cover roughly 55% of expenditure in Homescan.

Moreover, there are many brands and types of coffee and most coffee brands are rather small in market share. At such, I identify top brands as brands with at least 5%

of market share in either Keurig platform or ground coffee. For imputation purpose – top 95% brands in Keurig or top 85% in ground (Union).

Households also shop in many different chains, but most purchases are made in a couple of chains. For example, top 10 chains cover 61% of the K-Cup sales, and top 25 chains cover 80% (both in Homescan). For ground coffee, top 10 chains cover 45% of the sales and top 25 chains cover 67%. Many chains are also in RMS, in which case, require little imputation on price and availability given the rather homogeneity of good prices within a market and chain. However, chains only in HMS require imputation of price and availability of products based on the chain identity. At such, we select all RMS chains and chains in 25 top sales of K-Cup coffee. Such a criteria give us a list of 122 chains.

For the top 25 Homescan chains (some are in RMS), I impute the price as such based on the analysis in Hitsch et al. (2017). We found that most of the price variations can be explained by the chain and DMA dummies, and this is true for most chains. Also, prices are rather homogeneous within chains, which is the basis for price imputation. Price panel is constructed as DMA-Chain-Brand-Brand Characteristics (including roast level)-week pair. I then fill the price in the following order:

1. DMA-Chain-Brand-Brand Characteristics (not roast level) week average price
2. Region-Chain-Brand-Brand Characteristics (not roast level) week average price
3. Chain-Brand-Brand Characteristics (no roast level) week average price

At this point – 60% of price data is filled, and the remaining will be filled using 11 price regressions models. I first compute the average price various levels of geography, chain identity, time period, and product characteristics. This yield 12 price series with different levels of missing and accuracy. I then regress the price of the product on various combinations of these constructed average prices as well as product characteristics. I tried 11 different combinations. Then based on the goodness of fit of these 11 models, I predict the prices for the missing ones only. When I compare the imputed prices with prices in RMS, the correlation is 0.913.

APPENDIX F

KEURIG MACHINE ADOPTION IMPUTATION

F.1 Adoption Date Imputation

I face three challenges pertain to the imputation of household adoption of Keurig machine:

1. Households may be observed to purchase the portion packs many times, but never reported to have purchased the machine.

2. Even if the households report to purchase portions, the inter-purchase interval may be very long and "irregular" indicating they may forget to scan the portion pack purchases.

3. Households may drop out and reenter the panel for any given panel year.

For households who made purchases of both hardware and software, the association is only established if software purchase is made within -45 days and 180 days. Otherwise, the purchase of the machine could have been a gift, returned product, or an upgrade purchase. For software and hardware purchases not within the period, the association is not established.

To impute the adoption date, I first run LASSO regression of household characteristics on adoption date to consumption date for households whom I observe both hardware and software purchases. I then compared the predicted days to the actual days, but the model prediction is very poor. A quick inspection of the data shows that majority of the customers buy the machine buy K-Cups within 14 days. Therefore, I will flag the hardware acquisition date as the first day they buy the portion packs for households who are only observed to purchase K-Cups. For households who are not observed in previous panel year, but subsequently observed in current panel year, I build a churn in burn in period of 45 days to ensure the adoption is during that panel year. For households not present in the previous panel year, I don't impute the carrying date for households who made the first portion purchase within the 45 days of the year since it is very difficult to infer that they don't own the machine during the previous year. These households are excluded from the adoption estimation data but included in the overall coffee preference estimation data.

F.2 Keurig Machine Price Imputation

The imputation procedure is similar to that of coffee, and are only imputed for top 30 DMAs. However, one important distinction has to be made. There are several series of Keurig machines, and the purchases in Homescan is scant. As many households purchase the machine from department stores, which have low coverage in RMS. Therefore, the imputation is based on the scant data, and I only impute one price index for all series of Keurig machine. The adoption could be of any series, but all use the same price index. As different series of Keurig machines are frequently promoted together, the imputation captures the general pricing pattern of Keurig machines. The price coefficient may be attenuated due to measurement error, but the focus of the analysis is on the impact of utility gain on adoption behavior.

APPENDIX G

INVENTORY IMPUTATION

In Homescan data, households only report their purchases, not their consumption. However, households may respond to their inventory and adjust the purchase schedule. As such, I impute the inventory status based on average purchase/consumption rate, and I then use the imputed inventory to infer the probability of making a coffee purchase. The imputed inventory doesn't affect the product choice conditional on making a purchase. The inventory is imputed based on the following steps:

1. Compute the average consumption/purchase rate:

   (a) Obtain the total number of weeks the household exists in the panel before and after the Keurig machine adoption.

   (b) Calculate the numbers of servings purchased before and after the adoption.

   (c) Compute the average computation/purchase rate before and after the adoption.

2. Impute the initial inventory status based on the first purchase:

   (a) Obtain the first coffee purchase date of the household

   (b) Compute the initial inventory level as

   $$I_0 = \text{Average Consumption Rate} \times (\text{First Purchase Week - Initial Week})$$

   (c) If the first (First Purchase Week - Initial Week) ≥ 52, $I_t = 0$ for $t \leq$ First Purchase Week. This adjustment is done to avoid over imputation.

3. Compute inventory level:

$$I_t = \max(I_{t-1} + \text{Purchase Quantity}_t - \text{Average Consumption Rate}, 0).$$

The above equations gives the inventory level regardless of purchase types (K-Cups and ground coffee). For type-specific inventory level, if the current purchase type is different from last purchase type (e.g. current purchase is of K-Cups, and last purchase was of ground coffee),

$$I_t = \max(\text{Purchase Quantity}_t - \text{Average Consumption Rate}, 0).$$

I currently use type-specific inventory in the estimating purchase probability.

APPENDIX H
MEASURING CONCENTRATION IN PURCHASES

H.1 Concentration in Brand Purchased

To measure concentration, I constructed alternative measure using expenditure on the top purchased brand (C1), the top two purchased brands (C2), and the top three purchased brands (C3). Figure L.1 compare the C1, C2, and C3 for Keurig machine adopters before and after adoption. In this graph, we observe a notable shift from more concentration expenditure to a less concentrated expenditure distribution after the household's adoption of Keurig machine.

H.2 Concentration in Brand and Flavors Purchased

Figure L.2 compare the C1, C2, and C3 for Keurig machine adopters before and after adoption. In this graph, we observe similar shifts as for brands. However, we see much less concentrated purchase pattern compared with that of brands.

APPENDIX I

ROBUSTNESS TO LEARNING

Households may try a variety of K-Cup brands after their Keurig machine adoptions to learn about different brands, and their variety purchases may only be temporary. As such, I compute the household purchase concentration statistics two years after their adoption of the Keurig machine, and compare that with their purchase concentration before their Keurig machine adoption. Figure L.3 shows the HHI by time periods conditional on Keurig machine adoptions. The HHI distribution within 2 years of the adoption looks similar to that of 2 years after the adoption, though their purchases become a little more concentrated after 2 years (Panel 2 v. Panel 3). Overall, households seem to purchase more variety of brands even after 2 years of their adoptions (Panel 1 v. Panel 3). Therefore, if the learning process takes less than 2 years, learning about brands may not be a big contributor to the purchase of brand variety on the Keurig platform.

APPENDIX J

OUTSIDE OPTIONS

In this section, I show that various outside options used in the literature are not good choices in my setting. Inflation and household specific spending levels may confound my analysis of the outside option. Thus, I approach the problem in the following manner:

1. Keep only the households who are present in 2011, 2012 and 2013. Further, only keep households who hadn't adopted the Keurig machine or adopted in 2012.

2. Compute their annual spending levels in 2011 and 2013 on caffeinated drinks, drinks, grocery, and all store spending. The spending numbers exclude spending on the coffee category regardless of coffee type.

3. Compute the ratio of 2013 spending to 2011 spending for each household, which removes both overall price levels effects and household specific spending level effects.

Figure L.4 shows the histograms of the ratios of spending across caffeinated drinks, drinks, grocery, and all store spending by adoption status. Compared with reference group, adoption group spending distribution looks very similar without systematic shift. Thus, these outside options aren't good choices for an outside option of the coffee category.

APPENDIX K

COUNTERFACTUAL METRICS

Four metrics are of particular interest in the counterfactual scenarios:

- The expected adoption of the Keurig platform
- The expected total revenue loss/gain for the aftermarket goods
- The expected revenue loss/gain from GMCR-owned brands
- Welfare when different varieties are supplied

K.1 Adoption of the Keurig Platform

For the purpose of simulation, I assume the pricing path and distribution remain the same as before. The only thing changed is the removal of certain K-Cup brands from the household choice set. Because most people adopted after 2008, I treat the first week of 2008 as $t = 1$.

Conditional on the household not having yet adopted,

$$\Pr(y = 1|\iota = 0, P, \delta) = \frac{\exp\left(v_1(\iota = 0, P, \delta|\theta_a)\right)}{\sum_{i\in\{0,1\}} \exp\left(v_i(\iota = 0, P, \delta|\theta_a)\right)}.$$

Therefore, the probability of non-adoption for household i up to period t,

$$F_{it}(\iota = 0; P_i, \delta_i) = \prod_{s=1}^{t} \Pr(y_{is} = 0|\iota_{is} = 0, P_{is}, \delta_{is}).$$

Then, the population non-adoption rate assuming independent households is,

$$\bar{F}_t(\iota = 0; P, \delta) = \frac{1}{n}\sum_{it}^{i} F_{it}(\iota = 0|P_i, \delta_i).$$

The adoption rate at time t is then,

$$\bar{A}_t(P,\delta) = 1 - \bar{F}_t(\iota = 0; P, \delta).$$

This difference in the adoption rate at t for utility gain δ in the original case and $\tilde{\delta}$ in the counterfactual case give the change

$$\Delta A_t(P,\delta,\tilde{\delta}) = \bar{A}_t(P,\delta) - \bar{A}_t(P,\tilde{\delta}).$$

K.2 Revenue Loss/Gain

I'm interested in two measures of revenue loss/gain: (a) total revenue from K-Cups, and (b) revenue for GMCR owned K-Cup brands. For the purpose of computing revenue loss, I assume the household maintains the same shopping frequency. Therefore, the total revenue loss/gain can occur for two reasons: (a) the number of households with the Keurig machine, and (b) household coffee type substitutions. For GMCR revenue, the change can occur for one additional reason: households substituting from other K-Cup brands to GMCR-owned brands. The revenue loss and gain are computed without the licensing-fee scheme.[1]

In section K.1, I showed how the adoption rate can be computed. Given the adoption rate for each household $A_{it}(P_i, \delta_i)$ at time t and conditional on shopping occasion t, the household's expected expenditure on product k can be estimated by Monte Carlo simulation. For each Monte Carlo draw of random utility vector ε, and

1. The exact royalty fee structure is complex and confidential. The fees are individually contracted especially for large roasters. I called an alumnus of Booth who works at Starbucks. He explained that the fee structures are complex and strictly confidential, and thus he couldn't provide me with any details.

conditional on the household already having adopted the Keurig platform,

$$e^k(\varepsilon, \iota = 1) = \arg\max U_h\left(E - \left(\sum_{k \in G_{rt} \cup K_{rt}} e_k\right), e_1, \ldots, e_{|G_{rt} \cup K_{rt}|}\right).$$

If the household hasn't adopted the platform, the expenditure is 0 because of the consumption constraint. The expected expenditure on product k is then

$$\mathbb{E}(e^k) = A(P, \delta) \int e^k(\varepsilon) dF(\varepsilon).$$

Empirically, the above equation can be computed as,

$$\mathbb{E}(e^k) = A(P, \delta) \sum_{d=1}^{D} e^k_d(\varepsilon_d),$$

where D is the total number of Monte Carlo draws. The per-capita expected revenue of product k is computed by integrating over the distribution of households, and empirically it is computed as

$$\mathbb{E}(e^k) = \frac{1}{N} \sum_{h=1}^{N} \mathbb{E}(e^k_h).$$

Given product revenue, I aggregate all the GMCR-owned products to create the overall GMCR-owned brand revenue. Restricting the choice of households allows me to compute the expected revenue in various choice scenarios. The revenue loss/gain can be computed by comparing the revenue in different choice scenarios.

K.3 Welfare

For welfare analysis, I use Kim et al. (2002) approach by solving the compensating value (CV) through numerical optimization. For a given choice occasion and expenditure, the indirect utility is

$$\mathbf{V}^*(\mathcal{S}, E) = \max_{\{e_k\}, k \in \mathcal{S}} \psi_0 \left(E - \sum_{k \in \mathcal{S}} e_k \right) + \left\{ \sum_{k \in \mathcal{S}} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho},$$

and when the choice set is constrained to $\tilde{\mathcal{S}}$, the indirect utility function is

$$\mathbf{V}^*(\tilde{\mathcal{S}}, E) = \max_{\{e_k\}, k \in \tilde{\mathcal{S}}} \psi_0 \left(E - \sum_{k \in \tilde{\mathcal{S}}} e_k \right) + \left\{ \sum_{k \in \tilde{\mathcal{S}}} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho}.$$

Assuming the budget E is big enough and the outside good is always purchased, the increase of budget doesn't affect the utility from the inside good given the special structure of the utility function. Because the compensating value is estimated by solving the equation

$$\mathbf{V}^*(\mathcal{S}, E) = \mathbf{V}^*(\tilde{\mathcal{S}}, E + CV),$$

the CV can be simply written as

$$\begin{aligned} CV &= \left\{ \sum_{k \in \mathcal{S}} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k^*}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho} - \psi_0 \left(\sum_{k \in \mathcal{S}} e_k^* \right) \\ &\quad - \left\{ \sum_{k \in \tilde{\mathcal{S}}} \frac{\psi_k}{\alpha_k} \left[\left(\frac{e_k^\star}{p_k} + 1 \right)^{\alpha_k} - 1 \right] \right\}^{\rho} + \psi_0 \left(\sum_{k \in \tilde{\mathcal{S}}} e_k^\star \right), \end{aligned}$$

where e^* and $e^\star$ are the optimal expenditures. One potential catch is the random utility shock, which requires integration to create the expected CV. I estimate the expected CV using Monte Carlo integration. Conditional on the adoption and shopping occasion, household h's overall CV is computed as the sum of CV across shopping occasions

$$CV_h^{(i)} = \sum_{t=1}^{T_h} CV_{ht}^{(i)}.$$

To compare across households, I can use the percentage metric as proposed in Kim et al. (2002)

$$PCV_h^{(i)} = \frac{CV_h^{(i)}}{\sum_{t=1}^{T_h} E_{ht}}.$$

APPENDIX L

ADDITIONAL GRAPHS

Figure L.5 shows the scatter plot of MCMC draws of mean satiation parameter (α) for ground coffee and K-Cups.

Figure L.6 shows the scatter plot of MCMC draws of mean satiation parameter (α) for ground coffee and K-Cups.

Figure L.1. Concentration of Brands by Adoption Status

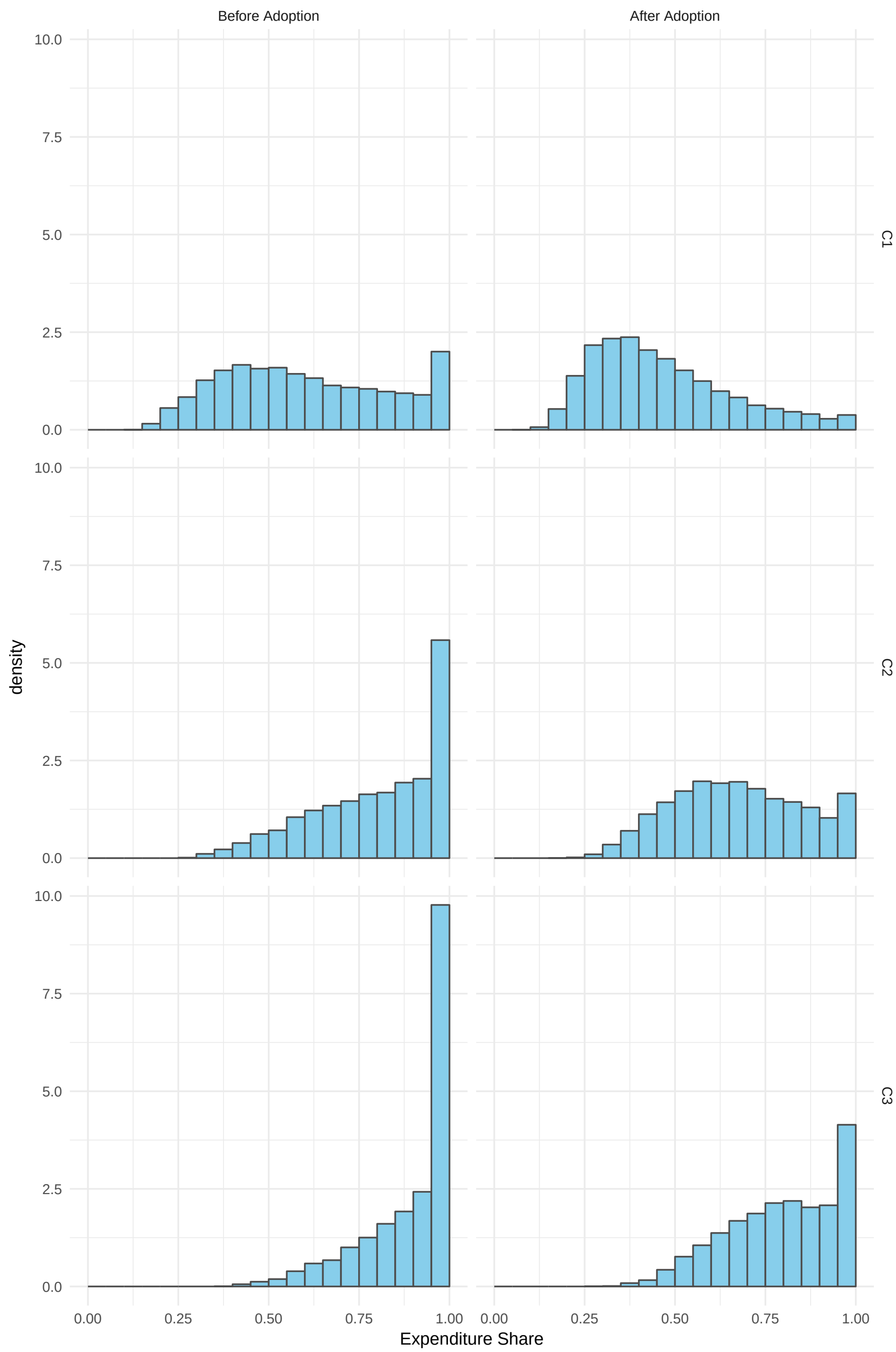


Figure L.2. Concentration of Brands and Flavors by Adoption Status

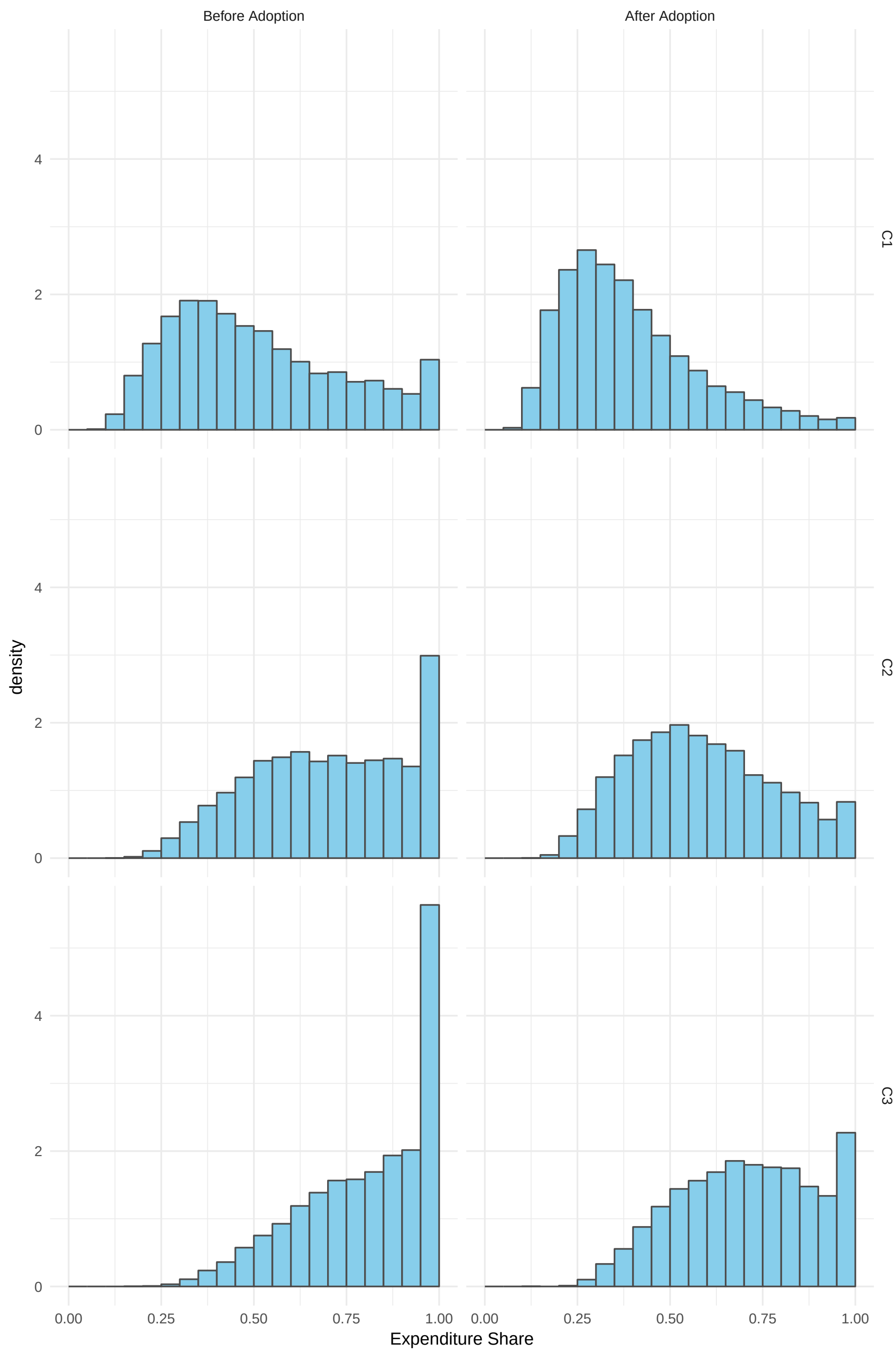


Figure L.3. Purchase HHI of Brands by Time Periods Conditional on Adoption

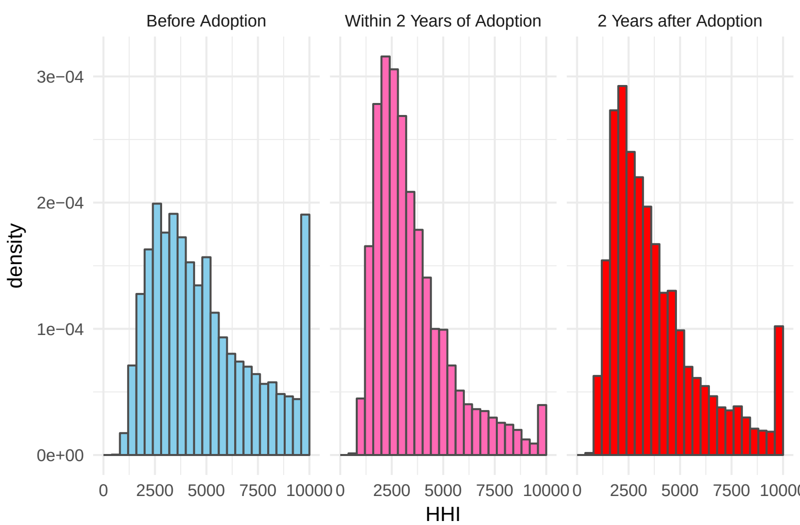


Figure L.4. Spending Ratios of 2013 to 2011 by Adoption Status

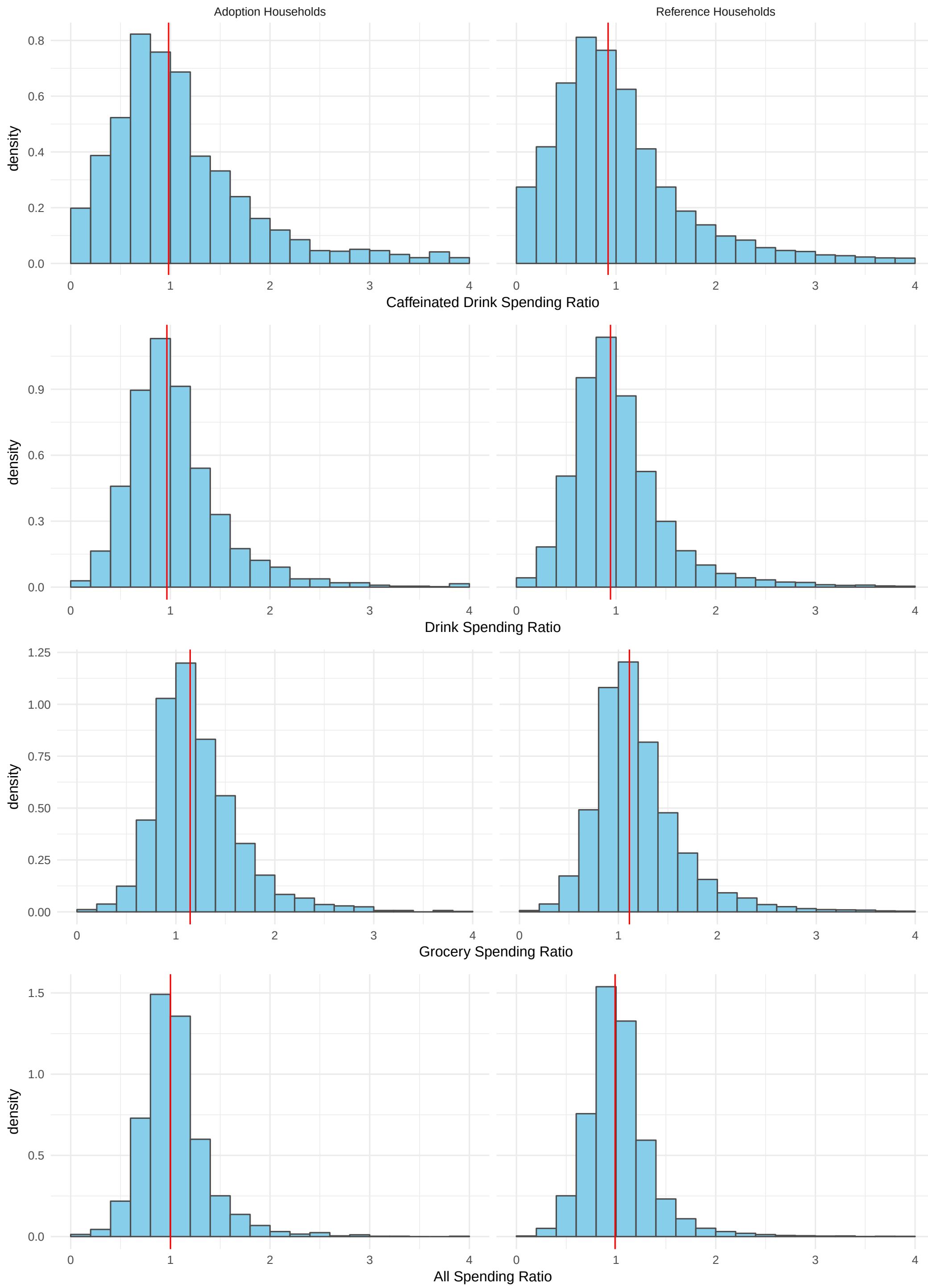


Figure L.5. Scatter Plot of MCMC Draws of Satiation Parameters (α)

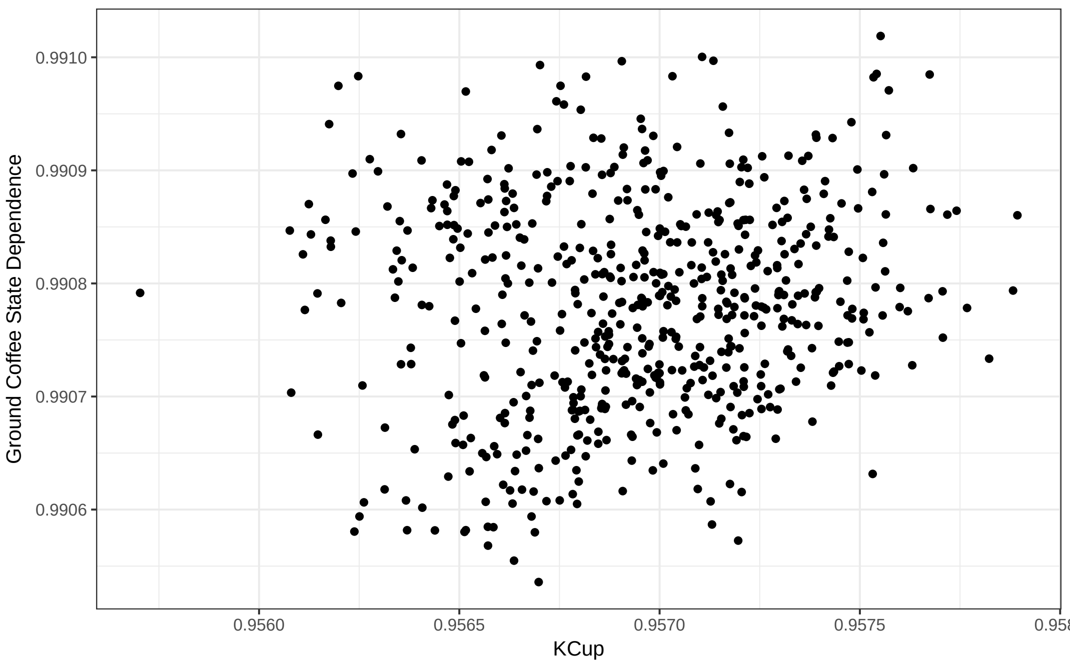


Figure L.6. Scatter Plot of MCMC Draws of State Dependence and K-Cup Adjustments