THE UNIVERSITY OF CHICAGO

CUSTOMER RETENTION UNDER IMPERFECT INFORMATION

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

YEWON KIM

CHICAGO, ILLINOIS

JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

ABSTRACT

I study why many firms face low retention rates among new customers. In particular, I examine whether customer churn at the firm level after a single product experience is solely driven by heterogeneous preferences or is affected by incomplete information about the products. I use a long panel of individual-level ticket purchases from a major U.S. symphony center for which 60% of first-time customers do not return after a single visit. The data exhibit patterns consistent with consumer learning and incomplete information about underlying values of concerts at the ticket purchase stage. Descriptive analyses show that imperfect information and learning spillover jointly cause customer attrition. First, many customers attend concerts with a low match value due to their incomplete information. Second, a low match value at the initial visit leads to high attrition rate at the symphony center level, which suggests that the initial visit experience creates strong learning spillovers by affecting a customer's expectations about all future concerts. To explore marketing strategies to reduce customer attrition, I develop a structural model that incorporates the learning spillovers and incomplete information. Through counterfactual analyses, I analyze both a policy that offers high-value concerts to first-time customers and a policy that offers targeted marketing to second-time customers after their initial visit. The results emphasize the importance of introductory marketing to new customers.

CHAPTER 1

INTRODUCTION

Many firms across different markets face low retention rates among new customers. Although a firm represents a large collection of diverse products in many cases, it is still commonly observed that more than 50% of consumers churn at the firm level after a single product experience in various markets.[1] Understanding the source of low retention rate after initial trials is critical for a firm to set optimal marketing policies. If the low retention rate is driven by preferences of fully-informed consumers who are well aware of the firm's entire product offerings, lowering price or changing product design would effectively increase retention. However, if the low retention rate is due to the imperfect information about the firm's available products, then retention would be increased via other marketing interventions like informative advertising or temporary price discounts that nudge marginal customers to explore more products.

I study the source of low retention among new customers in the context of classical music concerts. In particular, I examine whether customer churn at the firm-level is solely driven by consumer preferences or is also affected by imperfect information about available products in the firm. I use a panel of individual-level ticket purchases from a major U.S. symphony center for which 60% of first-time consumers do not return after a single visit at least for 4 years.[2] The data contain 13-year-long ticket purchases, and each observation records a unique consumer ID with the name of concert tickets purchased. Purchase data is combined with program catalogs that offer detailed information on the programming of concerts.

Instead of assuming consumers' imperfect information, I start my analysis by testing whether or not consumers have fully informed, fixed preferences for all available concerts. The test I develop extracts a testable prediction from a class of consumer learning models

---

1. "Blue Apron: Inside the box," *Data Points.*
"Ahead of IPO, Airbnb's consumer sales surpass most hotel brands," *Data Points.*
"Takeout takeover: Uber Eats now bigger than Grubhub in 15 major U.S. cities," *Data Points.*

2. The retention rate is calculated based on the customers whose travel distance is under 50 miles.

that has been widely used in the marketing literature. Under a standard learning model in which consumer belief about an unknown quality converges to the true quality with growing consumption experiences, experienced consumers' choices should reflect actual product qualities, and therefore inexperienced consumers' product choices should become more concentrated around the products chosen by experienced consumers as they learn about the underlying qualities over time. The data show that, as new customers become more experienced over time, their concert choices converge towards the choices made by a holdout sample of already experienced customers. This specific pattern of within-customer choice evolution over time is not explained under a model that assumes fully informed, fixed heterogeneous preferences. This test does not require a specific functional form of learning, such as normally distributed prior and signals, and therefore can be used to document reduced-form evidence of consumer learning.

Next, based on the assumption of consumer learning, I derive and estimate concert values (i.e., average preference for each concert perceived by the market) by inverting the choices of experienced consumers.[3] Observing the data pattern that rejects fixed preferences and supports consumer learning, I assume that experienced consumers are better informed about the underlying concert values at the purchase stage, and use their purchases to extract information about true concert values. The estimated measure of concert values passes several validity checks and allows me to document new consumers' lack of information about the concert values at the purchase stage. It also allows me to trace the learning path of consumers without assuming a specific learning rule by observing how their choices evolve in terms of concert values over time.

Descriptive analyses show how imperfect information causes customer attrition jointly with learning spillovers. First, many new customers attend low-average-match-value concerts due to their incomplete information at the initial purchase stage. The arrival rate to a concert among first-time visitors is random with respect to its underlying concert value revealed by

_____

3. Concert values can contain vertical quality, horizontal match value, or both.

experienced consumers' choices. Second, a low match value at the initial concert visit has a significant impact on the probability of subsequent churn at the symphony center level, even after controlling for a rich set of confounding factors and potential heterogeneous preferences among the first-time customers. This indicates a strong spillover effect of a single concert experience on a customer's expectations about all future concerts. In other words, consumers treat a single experience as representative of what the symphony center offers, although the concert they consumed lies on the low tail of match value distribution. The pattern of strong spillovers from the initial product experience to consumer perception of the firm's entire product portfolio has not been formally addressed in empirical marketing literature, although consumer psychology theories support the significant impact of initial experience in various contexts.

To explore how the symphony center can reduce customer attrition under this mechanism, I propose a new structural model of consumer learning. The model allows for flexible patterns of learning spillover in a computationally tractable way. In my model, customers extrapolate their past experiences to predict the value of other untried concerts by taking the weighted average of the past experiences. The model reflects several data patterns that are not fully captured by traditional consumer learning models. For example, brand-level learning models assume that each signal about a brand is randomly drawn whenever consumers make purchases, whereas in my data set I observe that the signal about the symphony center is endogenously selected by consumers via concert choices. Attribute-level learning models assume that information about different features can be obtained via consumption of those features, whereas in my data consumers become more likely to purchase high-value concerts as they become more experienced even when they face new concert features they have not tried before. To reflect these patterns, I allow consumers to acquire information about upcoming concerts through an additional channel besides consumption, which I call "search". Here, search does not mean online search activities but refers to any information acquisition behavior, ranging from paying attention to the content of promotional materials to reading

product reviews. In my model, consumer decision to engage in search is determined by how pleasant the previous concert experiences are, which creates an incremental impact of prior concert experiences on subsequent ticket purchase behavior. Specifically, a satisfying prior experience not only raises the likelihood of returning to the firm but also increases the probability of purchasing high-value products in the next period due to increased search, which generates room for endogenously selected signals about the firm.

The key parameters of the estimated structural model reflect 1) how far consumers generalize the information from a single concert experience to all the other concerts, and 2) how consumers obtain additional information on concert values via search based on their previous visit experiences. The estimated parameters suggest high experience spillovers and low search activities in the first few visits, explaining the lasting impact of few consumption experiences on customer retention under imperfect information.

Counterfactual analyses offer rich managerial implications both on customer acquisition and retention efforts. First, the results highlight the importance of introductory marketing that steers new customers towards better first-time experience. Simulation using structural parameters shows that even 70% discount offered on the second visit is not sufficient to match the effect of high-value initial experience on the number of return visits. Second, the results suggest that targeting retention campaigns based on customers' initial concert visits can increase the return on retention efforts. Retention efforts can be futile if they target customers who visited bottom-value concerts because of the strong adverse spillovers from the initial visit. Similarly, the same efforts can also generate low incremental revenue if they target customers who visited top-value concerts, since those customers already have higher incentives to return due to the strong positive spillovers from the initial visit.

Counterfactual exercises also emphasize the potential trade-offs the symphony center would face when making marketing decisions given the specific learning behavior. For example, enriching concert variety can have two opposing effects on the symphony center's profit. On one hand, it can increase the profit by raising the arrival rate of visitors and satisfying

the tastes of broader audience. On the other hand, it can decrease the profit by increasing the probability of consumer-concert mismatches and subsequent churn. Counterfactual indicates that reducing concert variety by removing concerts with bottom 1% concert values can raise both ticket revenue and the average number of visits, underlining the negative impact of a large choice set on customer retention due to information problem. Similarly, price promotions put on low-value concerts can result in two diverging impacts. While it can increase the profit by selling remaining seats, it can also decrease the profit by nudging first-time visitors to try low-value concerts and to lose interests in all the upcoming concerts. The results show that giving discounts on low-value concerts to first-time visitors generates lower number of visits and lower ticket revenue than giving the same amount of discounts on high-value concerts. These trade-offs suggest that a firm, when setting its marketing strategies, should explicitly consider that customers may rely on the very first experience to determine their subsequent relationship with the firm given their limited information about its products. The findings also imply that a firm should pay more attention to informing its new customers instead of focusing only on its loyal return customers ([50]).

This paper complements the existing literature on the drivers of customer churn by applying consumer learning models. Customer churn or retention as a topic has been extensively discussed in marketing literature ([52], [31], [33], [65], [3], [6], [17], [39], [58]). However, a large amount of effort has been put to predict when customers churn, and there have been surprisingly few studies that explain why customers do not return ([2]). Moreover, to my knowledge, there is no empirical research that looks at why such a high number of churn events take place at the early consumption stage, although most statistical models that predict churn fully take this pattern into account when fitting the data. Using micro-founded consumer utility model, I view churn as an explicit outcome of consumer learning, which offers useful insights on how to design marketing interventions to prevent churn ([39], [58], [43], [4], [5]). In addition to increasing a firm's profit, these interventions can increase consumer surplus by facilitating consumer learning that otherwise might have stopped, as literature in

5

various disciplines implies the potential welfare loss caused by incomplete information ([42], [59], [25], [38]).

The paper complements the literature on consumer learning. It shows reduced-form evidence of learning by testing different predictions under the model with and without consumer learning. It proposes a new framework that allows for flexible learning spillovers as well as incorporates an additional endogenous information acquisition activity besides consumption. It also illustrates how the initial consumption experience - which is assumed to be "debiased" with more purchases in Bayesian learning literature - can have a much bigger weight on people's beliefs and determine the entire subsequent consumption path, as documented in consumer psychology literature ([61], [41]) and few empirical studies ([7], [35]). More broadly, the paper discusses the gap between what consumers learn about and which signals they use. Facing rapidly changing choice sets, consumers extract information about a firm's entire product offerings by sampling one or few of its diverse products, rather than learning about a relatively homogeneous group of products by trying one of them. This implies that the signals consumers receive and the construct they learn about no longer align perfectly, and that experience spillovers via correlated learning may take place very strongly and broadly ([28], [23], [56], [60], [18], [20]).

Finally, this paper proposes a new approach to identify unobserved product qualities from purchase data only. Although the idea of using choices by individuals outside the group of interest has been applied in prior research to control for unobservables ([44], [15]), it is new to my knowledge to recover product qualities using the choices of a consumer subgroup based on the prediction of consumer learning models. Estimating product qualities using the revealed preferences of informed consumers can be useful for several reasons. First, it is hard to find survey data on product experiences that can be used with matching purchase data. Second, survey data may not be informative of consumers' true preferences in certain cases. Prior literature raises concerns about various types of survey bias ([45]), and industries also question the validity of the survey results based on their low predictive power ([47]). The

proposed methodology would allow researchers to control for unobservable product values without an endeavor to find additional data sets.

The rest of the paper proceeds as follows. Chapter 2 discusses the related literature. Chapter 3 introduces the general setting of interest and the specific empirical context of the paper. Chapter 4 presents reduced-form evidence suggestive of consumer learning and the identification of concert values from purchase data only. Chapter 5 shows descriptive evidence of incomplete information about concert values and its impact on customer attrition at the symphony center level, and Chapter 6 proposes a framework of consumer learning that justifies the data patterns. Chapter 7 and 8 discuss the model specification and estimation, and Chapter 9 reports the results. Chapter 10 discusses counterfactual analyses, and Chapter 11 concludes.

CHAPTER 2

LITERATURE REVIEW

In this chapter, I explain how this paper is related to the literature on customer attrition and consumer learning.

2.1 Customer attrition

Customer attrition has been an important outcome of interest for both industries and academic literature ([9]). Research on customer attrition has studied two different settings: a contractual setting in which consumer departure is observable by a firm (e.g., not renewing subscription) and a noncontractual setting in which the timing of customer churn cannot be clearly labeled (e.g., not making another purchase). In both settings, many of the extant works focus on fitting a statistical model to predict churn using variables ranging from consumer demographics to consumption history ([52], [31], [33], [65], [3], [6]). Much less discussion has been made on understanding the causal drivers of customer churn ([10], [12], [13], [58], [2]). Moreover, although many models that predict churn reflect the commonly observed pattern of high churn rates at the early stage of consumption, there is no empirical study that focuses on why churn rate is higher among the first-time customers or whether there can be any improvement in retention at the early consumption stage.

This paper is related to the empirical works on customer attrition in a noncontractual setting. However, the main focus of this paper is not to predict when consumers churn but to understand why they churn. Using a micro-founded consumer utility model, I complement the existing literature by explaining customer churn as an explicit outcome of consumer learning, which offers useful insights on how to design marketing interventions based on customers' needs in their learning process ([39]). Also, the paper suggests that the high attrition rate among new customers should not be assumed to be an uncontrollable feature of the market but be viewed as a measure to be improved.

2.2 Bayesian consumer learning models

Researchers have developed different models to study how consumers learn about the undering-ing qualities of experience goods via purchases ([42]). Most learning models assume Bayesian consumers; consumers purchase a product that maximizes their expected utility given their prior information, and update their beliefs after receiving signals using Bayes' rule. Models vary in the length of time horizon that consumers consider when making decisions (myopic [[22], [21], See [57] for review] vs. forward-looking [[29], See [19] for review]). Models also vary in the scope of information spillover. Information from one purchase occasion can have spillovers to different categories, brands, or product attributes ([28], [23], [56], [60], [18], [20]).

Acknowledging the limitations of strong functional form assumptions in traditional Bayesian learning models, researchers have extended the modeling framework to incorporate how consumers can potentially deviate from full Bayesian updating. This framework, sometimes called Quasi-Bayesian learning model ([27], [46]), allows consumers to revise signals or prior beliefs in a subjective manner. ([34], [16])

Despite the richness of modeling approaches, empirical literature on consumer learning faces several challenges. First, most papers that propose intricate learning models do not examine whether or not learning actually drives consumer purchases prior to modeling. In many cases, identification of learning comes from functional form specifications (e.g., normally distributed prior and signals) ([19]), and only few papers document the evidence of (no) learning before developing a full structural model ([21], [26]). Modeling consumer learning based on the assumption of specific learning rule makes the results less interpretable, since it is hard to disentangle learning from preference heterogeneity in empirical settings ([54]). Second, most papers ignore the possibility that brand-level learning is based on a product-level signal, or that the product-level signal is an endogenous outcome of consumer purchase decisions. Prior research on learning assumes that a consumer receives a random signal about a brand from any product purchase, not allowing the case in which each signal

she gets is not fully random depending on which product she chooses. Finally, although consumers do not rely only on consumption experiences to obtain information about products or categories, there has been no empirical paper that models how consumption-based and non-consumption-based learning activities (e.g., search) jointly affect consumer purchase decisions.

This paper extends the existing literature in several ways. First, I show reduced-form evidence of learning by deriving a testable prediction from a class of learning models that are widely used in the literature. Although this evidence is based on a set of assumptions and therefore is not truly model-free, it adds strong support to the existence of incomplete information at the purchase stage followed by consumer learning over time. Second, I model that consumers choose a signal (product) based on their expected utility instead of randomly drawing one, and allow such signals from product experiences to affect the perception of a firm's entire product portfolio in a tractable way. Third, I allow consumers to get additional information from an activity besides consumption, which I call search. In my model, decision to search is endogenously determined by learning from previous consumption experiences.

2.3   Consumer judgment given limited information

Consumer psychology literature suggests that consumers' initial judgment may not only be more extreme ([51]) but also be harder to be adjusted by subsequent stimuli if the judgment is made with limited information ([61], [41]). The paper by [51] shows that "judgment may be overly favorable and confident when the limited information is positive in valence, whereas the judgment may be overly unfavorable and confident when the limited information is negative in valence." This suggests that, when inexperienced consumers have negative consumption experience from a firm's product with limited information about the firm, their judgment about the firm might be more negative and confident than the judgment made by experienced consumers with more information. In the language of Bayesian learning, this finding suggests that inexperienced consumers may interpret the signal to be more extreme

10

with lower variance than experienced consumers would.

The lasting impact of initial signals can also be explained by anchoring and adjustment heuristics ([61]) or the initial judgment effect ([41]). Anchoring followed by insufficient adjustments suggests that different initial signals can result in different estimates about the same value, which are biased towards the starting points ([61]). The initial judgment effect, first documented in the context of social judgments ([64]), indicates that initial judgments of a target affect subsequent memory-based judgments of the same target ([41]). Memory-based judgments based on the initial judgments are also correlated with people's interest in seeking more information about the target product ([41]). This implies that initial negative signal can block further information acquisition, which makes the impact of the negative signal on subsequent purchase behavior even stronger.

CHAPTER 3
SETTING

Chapter 3 introduces the general setting of interest and the specific empirical context of the paper. In particular, it discusses how imperfect information can be a driver of customer attrition in a market with multiple sources of incomplete information.

3.1   General setting - A market with imperfect information

The paper focuses on the market with two sources of incomplete information. The first source is that there is a component of product value (vertical qualities, match values, or both) that is not realized until the product is consumed. Examples include any markets with experience goods ([42]). The second source is that, given the size of choice set, consumers do not have full information on the range of product values that a firm offers. Examples include markets for clothing items, furniture, stationary products, and food items (e.g,. cereals), in all of which each firm provides more than a handful of choice alternatives. The second type of incomplete information suggests that, even if product values can be fully learned before consumption, consumers are unlikely to be aware of a firm's entire set of product values unless they exert significant efforts. For example, a consumer has to spend minutes or even hours to look through the catalog of all Nike sneakers which consists of more than 300 different styles.

Consumers engage in different information acquisition activities to reduce each type of information gap. To learn about the underlying value of a given product, consumers make purchases and realize the true product value via consumption. Facing a large set of alternatives offered by a single firm, consumers engage in information-seeking activity before purchase to learn what product values are available in addition to what product value each item has. I call this information channel "search." Here, search does not mean any specific online search activity but refers to generic information acquisition behavior besides consumption, ranging from paying attention to product catalogs to reading online reviews.

Each of the two information acquisition activities - search and consumption - results in a signal that consumers use to update their beliefs about the products of interest. In particular, beliefs or predicted consumption benefit created by prior consumption experiences determine the amount of subsequent search activities. For example, if the past product experience was positive, then consumers would search more products of the same firm in the next shopping period, which increases the likelihood of finding a high-value product from the firm's listing. In contrast, if the past product experience gave low satisfaction, consumers may turn off their incentives to search the firm's other untried items even if there exists high-value products they would buy had they searched.
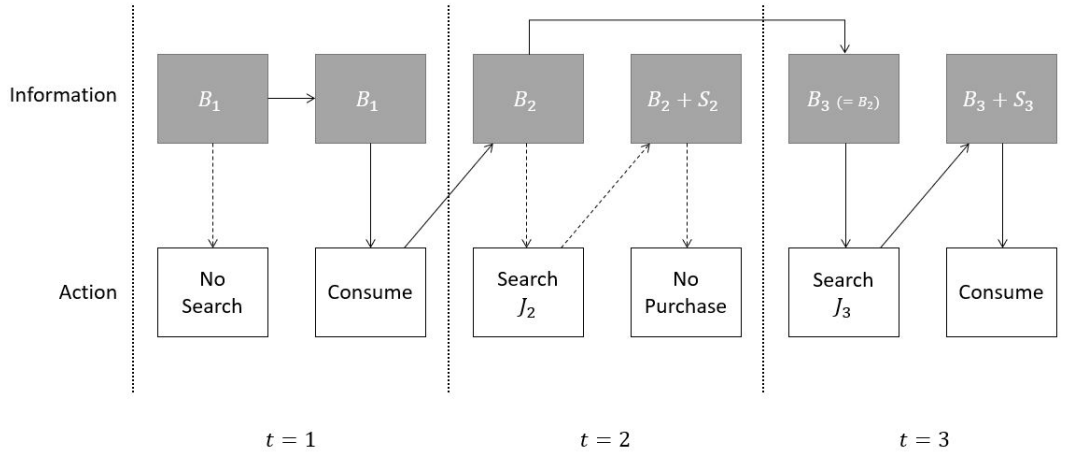
Figure 3.1: Consumer information acquisition process

Figure 3.1 illustrates how prior product experiences shape search and purchase decisions given the imperfect information. In each time period, before making a purchase decision, a consumer can engage in search to obtain specific information about items available at the moment (J_t). Her decision to search is based on the trade-off between expected gain and cost from search. Expected benefit from search depends on the perceived mean and variance of the untried product values, both of which are updated from the past consumption experiences.

Let B_t denote the belief about product values at time t that has been obtained from the previous product experiences. Let S_t denote the information on currently available product values obtained from search. At the search stage, a consumer decides whether or not to

search for more product-specific information (S_t) based on her belief carried over from her prior experiences (B_t). At the purchase stage, her information set is $B_t + S_t$ if she has engaged in search, and B_t if she has not.

In this framework, information from prior consumption experiences (B_{t-1}) affects a consumer's decision in the next period two ways. First, it directly enters the information set (B_t) based on which a purchase decision is made. Second, it affects the amount of additional information obtained through search (S_t), which also enters the information set for purchase. These direct and indirect effects of the experienced product values can generate a snowball effect on customer retention. Positive consumption experience not only directly raises the probability of returning to the firm but also raises the probability of purchasing a high-value item from the same firm due to the incremental information from search. Similarly, negative consumption experience not only directly decreases the probability of returning to the same firm but also decreases the probability of search, which deters consumers from finding the "right" products from the firm even if the highest-value product is indeed offered by the same firm.

In summary, in any market where there is imperfect information on individual product values and on the range of product values offered by a firm, a single product experience can exert a lasting influence on customer-firm relationship, even leading to churn, by affecting subsequent search and purchase intentions. Next, I introduce the specific empirical setting I study to address this effect.

3.2   Empirical setting - A market for classical music concerts

I study how imperfect information drives customer attrition in the case of a major U.S. symphony center. The symphony center hosts about 120 unique concerts every year, each of which can be viewed as a product (concert) offered by a firm (symphony center). The data set is an individual-level panel of ticket purchases for 13 fiscal years. Detailed information on each concert is extracted from program catalogs, which include but are not limited to

pieces performed, performers, soloists, and solo instruments.

The market for classical music concerts is a representative example of the general setting described in the previous section. First, concerts have both search quality that can be learned before purchase (e.g., how famous a soloist is) and experiential quality that can only be realized after consumption (e.g., how enjoyable a live performance of the specific soloist is). Second, given the large set of available concerts with various programs, consumers cannot visit every concert but necessarily rely on a small number of prior concert experiences to decide their subsequent search and ticket purchase decisions. Third, 60% of the first-time visitors do not return to the symphony center after a single visit at least for 4 years, which raises the question of why such a low retention rate is observed and how customer attrition can be managed via marketing interventions. Fourth, most promotional materials delivered to customers focus on general information on upcoming concerts instead of highlighting specific concerts only. This alleviates concerns about endogenous product choices of experienced consumers due to certain marketing activities.

3.2.1   Purchase data and consumer demographics

Table 3.1 reports descriptive statistics of the ticket purchase data. The data set contains purchases from FY2005 to FY2018, but for the analysis I use purchases from FY2009 to FY2015 to alleviate potential biases from left truncation. Given 4 years of burn-in and burn-out periods, new customers are identified as those who have not visited the symphony center at least in the past 4 years. Similarly, customer churn is defined as an occasion in which a customer does not return to the symphony center at least in the next four years. About 150,000 purchases are made each year by about 48,000 unique consumers. There are approximately 120 unique concerts held each year, and many of the concerts are performed more than once which results in a higher number of total concerts per year.

The median number of total ticket orders made by an individual customer over 7 years is 1 (Table 3.1(b)), although the distribution has a long right tail. Median interpurchase time

conditional on staying after the first visit is 44 days. Although the maximum number of days between two purchases is 2449 days($\approx$ 6.7 years), cases in which the interpurchase time is longer than 3 years is only 0.5% of the data, which lessens concerns for left truncation bias when labeling customer entrance or departure.

Price per ticket ranges from $0 to $750, and there is a significant amount of within-individual variance in prices paid across different concerts (Table 3.1(b)). These statistics are based on the actual prices paid by customers and not on the list prices offered by the symphony center. A major portion of variations in the list prices comes from different seat locations within a concert (See the last column of Table 3.2).

Other information on each ticket purchase observation include exact seat locations, number of tickets ordered, ticket sales channel (e.g., box office, online purchase), price promotion, ticket order date, performance date and time, and whether tickets are purchased as a bundle or as a single ticket. Ticket bundles consist of 3 to 4 individual concert tickets offered at a discounted rate. There are 70 to 120 different ticket bundles that are offered every season including a 'build-your-own' customized bundling option, which ensures flexibility in ticket choices even when customers decide to buy in bundles.

Consumer demographics are collected based on individual-level zip code information. Both per-capita income and distance to the symphony center show wide dispersion. Travel distance is used to distinguish local consumers from non-local visitors.

3.2.2   Concert features

Concert features are scraped from the text data of program catalogs provided by the symphony center. Ticket purchase data also contains certain information about individual concerts, such as different categories each concert belongs to.

There are 14 categories of concerts grouped by the symphony center. These categories vary in a list of features including the age of pieces performed (e.g., contemporary, classical), the genres of pieces performed (e.g., Jazz, movie sound tracks, classical pieces), musical com-

16

Table 3.1: Descriptive statistics

(a) By fiscal year

Fiscal Year	Purchases	Unique visitors	Unique concerts	Total concerts
2009	154,937	45,305	116	186
2010	163,058	48,112	121	200
2011	155,634	47,925	119	191
2012	157,190	49,428	118	187
2013	154,222	48,848	114	192
2014	153,046	49,898	116	196
2015	153,211	50,775	120	198

(b) By individual consumer

	Min	1Q	Median	Mean	3Q	Max
Total orders	1	1	1	2.8	2	294
Days between purchases (Among those with total orders >1)	1	21	44	109	118	2449
Price paid per ticket	0	29	47	55	77	350
Within-individual SD of the price paid	0	4	13	17	24	215
Per-capita income (zip-code level)	894	28,824	38,466	44,059	55,034	200,208
Distance (zip-code level)	0	6	17	145	39	4,310

(c) By individual consumer (within 30 miles; 61.3% of the total customers)

	Min	1Q	Median	Mean	3Q	Max
Total orders	1	1	1	3.35	2	294
#Days between visits (Among those with total orders >1)	1	21	49	144.8	134	2449
Price paid per ticket	0	20	40	46	64	350
Within-individual SD of the price paid	0	8	14	17	23	215
Per-capita income (zip-code level)	11,463	31,340	41,825	48,650	63,820	135,807
Distance (zip-code level)	0.0	3.39	8.05	10.70	16.85	29.94

(d) Churn rate at each visit (Among visitors who live within 20 miles)

17
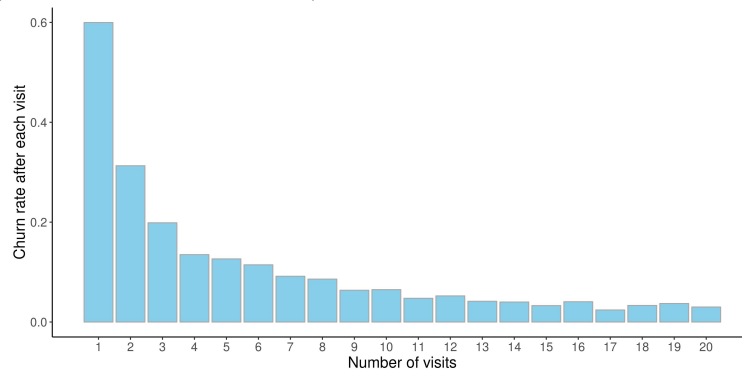
position (e.g., chamber, orchestra, solo), ambiance (e.g., casual and experimental, traditional and classic), whether the guest performers are invited, whether student artists perform, and the target audience (e.g., family-friendly). Each category has different baseline price, which is one of the two major sources of variations in price paid other than seat locations (Table 3.2).

"Specials" is a composite category that features performances by guest musicians including outdoor concerts, non-western music genres (e.g., Middle Eastern, Asian), non-classical music genres (e.g., Jazz), and holiday concerts (e.g., Christmas, Lunar New Year). To control for content diversity within this specific category, I manually put these concerts into existing categories if available and also create a separate category for Non-western/non-classical music.

There are more than 1000 different composers, conductors, orchestras, soloists, and solo instruments appearing in the data set. Including only those features that appear more than once in the entire data set reduces the dimension to 615. Program features used for analysis are summarized in Table 3.3.

Table 3.2: Category-level price (in $)

| Category | Unique concerts | All | | Within-concert |
		Mean	SD	Mean SD
Main	253	65.19	41.46	40.70
Guest Piano	47	33.92	18.66	15.18
Guest Chamber	29	50.17	28.98	20.33
Movies	21	54.90	24.12	23.16
Jazz	61	44.77	20.35	15.81
Casual classic	25	45.40	17.51	17.26
Specials	112	53.40	41.08	21.85
Casual fusion	22	44.51	27.10	25.57
Emerging professionals	43	0.00	0.00	0.00
Guest contemporary	23	13.77	6.46	6.23
Guest orchestra	23	54.78	33.81	31.82
Chamber	44	9.12	10.83	2.71
Family	23	19.83	10.29	10.16
Emerging professionals, fusion	8	9.12	7.55	1.41

Although a large portion of across-concert price variations comes from the difference in

Table 3.3: Program features

	Number of levels
Category (defined by the symphony center)	14
Genre (added by researcher based on the categories)	15
Composer	139
Conductor	119
Movement(Era)	7
Solo instruments	30
Country of origin	22
Orchestra	21
Solo artist	263
Total	615

baseline price across categories or genres, there is still within-genre across-concert variation in price. To capture these price differences across concerts, I compute the average price of the seats in Main Floor for each concert and use it as the concert price.[1] Table 3.4 reports the summary statistics.

Table 3.4: Concert-level price (in $)

Min	1Q	Median	Mean	3Q	Max
0.00	41.97	58.78	54.11	68.59	165.57

In the next chapter, I show the reduced-form evidence of imperfect information and learning about the underlying product values in the market for classical music concerts.

---

1. I only use the observations that entail full price to calculate the concert prices.

CHAPTER 4

A TEST FOR IMPERFECT INFORMATION

To study whether imperfect information causes customer attrition, I first develop a test that shows whether consumers have fully informed, time-invariant preferences for concerts at the purchase stage. This step of investigating whether consumers do face incomplete information in the context of interest is missing in many empirical studies that build rich learning models. The test contributes to the literature by introducing a way to check whether or not a data set shows a consistent pattern with imperfect information and therefore is suitable for studying consumer learning.

The test looks at the stability or time-invariance of consumer beliefs about product values over their purchase occasions. Here, product values can represent vertical quality or horizontal match value or both. If consumer beliefs about product values stay unchanged over time under the model of fully informed time-invariant preferences, then the choices of inexperienced consumers should not resemble those of already experienced consumers over time as the inexperienced consumers make more purchases. If the choices by these two groups of customers become more similar over time, then it implies that consumers' preferences, or beliefs about product values, changes as they make purchases. The pattern of changing beliefs about product values rejects the model of fully informed fixed preferences and supports the existence of incomplete information at the initial purchase stage.

4.1   A test for stability of consumer beliefs

I use the following two assumptions to construct hypotheses that test the stability of consumer beliefs:

· *A1*: True product value is realized upon consumption (Nelson 1970).

· *A2*: A consumer's choice probability is a function of a time-invariant utility component (α), her belief about the underlying product value (Q), and a mean-zero random utility

component (ϵ) which are additively separable. Random utility component is orthogonal to the belief about product value, and its distribution is known to researchers. That is, consumer i's predicted consumption utility from consuming product j at n-th visit (purchase occasion) is

$$u_{ijn} = \alpha_i + Q_{ijn} + \epsilon_{ijn}$$

and her choice probability is

$$s_{ijn} = \mathscr{S}(\alpha_i, Q_{ijn}).$$

$\mathscr{S}(\alpha_i, Q_{ijn})$ is 1) everywhere differentiable w.r.t. Q_{ijn}, and 2) $\frac{\partial \mathscr{S}_j}{\partial Q_{ijn}} > 0$ & $\frac{\partial \mathscr{S}_j}{\partial Q_{ikn}} < 0$ $\forall k \neq j$ ([8]).

Let $\mathcal{E} = \{i | n \geq \bar{n}\}$ denote a sample group of experienced consumers whose number of past visits (n) is greater than or equal to $\bar{n}$. Similarly, $\mathcal{I} = \{i | n < \bar{n}\}$ denote a *separate* sample of inexperienced consumers. $\mathcal{E} \cap \mathcal{I} = \emptyset$.

Remark 1. Suppose A1-A2 hold. Then, under the model of fixed preferences with perfect information (in which consumer belief is stable: $Q_{ijn} = Q_{ij}$ $\forall n$), the correlation between $\mathcal{I}$'s and $\mathcal{E}$'s product choices does not change as $\mathcal{I}$ make more purchases. In other words, if the correlation between the two sets of observed choices changes over $\mathcal{I}$'s purchase occasions, the beliefs are not stable.

Sketch of proof: Here, I show that (A: stable consumer belief) $\Rightarrow$ (B: no time-varying correlation between $\mathcal{I}$'s and $\mathcal{E}$'s choices). The second part of Remark 1 is $\neg B \Rightarrow \neg A$.

Let $\boldsymbol{s}_{\mathcal{E}}$ and $\boldsymbol{s}_{\mathcal{I}}$ denote a $J \times 1$ vector of observed product choice shares of experienced and inexperienced consumer groups respectively. $\boldsymbol{Q}_{\mathcal{E}}$ and $\boldsymbol{Q}_{\mathcal{I}}$ denote a $J \times 1$ vector of average beliefs on product values of experienced and inexperienced consumers, and $\boldsymbol{\epsilon}$ denotes a $J \times 1$ vector of random utility components. α_k denotes a time-invariant utility component of consumer group k. $n_{\mathcal{I}}$ represents the number of visits $\mathcal{I}$ have made.

21

If consumer beliefs about product values are stable, neither $\boldsymbol{Q_I}$ nor $\boldsymbol{Q_E}$ change over purchase occasions, which implies the following:

$$\boldsymbol{s_I} = \mathscr{S}(\alpha_I, \boldsymbol{Q_I})$$

$$\boldsymbol{s_E} = \mathscr{S}(\alpha_E, \boldsymbol{Q_E})$$

$$\frac{\partial \; corr(\boldsymbol{Q_I}, \boldsymbol{Q_E})}{\partial n_I} = \frac{\partial \; corr(\alpha_I, \alpha_E)}{\partial n_I} = 0$$

$$\Rightarrow \frac{\partial \; corr(\boldsymbol{s_I}, \boldsymbol{s_E})}{\partial n_I} = 0 \qquad \qquad \square$$

Next, I illustrate the prediction of standard learning models about the same construct: how the correlation between the two consumer groups' product choices would change over time under the model of standard consumer learning. I introduce an additional assumption that holds for a large class of learning models, including standard Bayesian learning models.

· *A3*: Under the model of consumer learning, consumer belief about product values converges to the true product values (Q_j^*) as consumers accumulate more consumption experiences. That is,

$$Q_{ijn} \to Q_j^* \text{ as } n \to \infty.$$

Remark 2. Suppose A1-A3 hold. Then, the correlation between $\mathcal{I}$'s and $\mathcal{E}$'s product choices increases with $\mathcal{I}$'s number of purchase occasions $n < \bar{n}$.

Sketch of proof: (I use the same notation from Remark 1).

Under the model of learning that satisfies A1-A3, $\boldsymbol{Q_I}$ converges towards a vector of true

22

product values $\boldsymbol{Q^*}$ as $\mathcal{I}$ accumulates more experiences (n).

$$\boldsymbol{Q}_{\mathcal{I}_n} \to \boldsymbol{Q^*} \approx \boldsymbol{Q}_{\mathcal{E}} \quad \text{as} \ \ n \to \infty \quad (\because A3)$$

$$\boldsymbol{s}_{\mathcal{I}} = \mathscr{S}(\alpha_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{I}n})$$

$$\boldsymbol{s}_{\mathcal{E}} = \mathscr{S}(\alpha_{\mathcal{E}}, \boldsymbol{Q}_{\mathcal{E}}) \approx \mathscr{S}(\alpha_{\mathcal{E}}, \boldsymbol{Q^*})$$

$$\frac{\partial \ corr(\alpha_{\mathcal{I}}, \alpha_{\mathcal{E}})}{\partial n_{\mathcal{I}}} = 0, \quad \frac{\partial \ corr(\boldsymbol{Q}_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{E}})}{\partial n_{\mathcal{I}}} > 0$$

$$\Rightarrow \frac{\partial \ corr(\boldsymbol{s}_{\mathcal{I}}, \boldsymbol{s}_{\mathcal{E}})}{\partial n_{\mathcal{I}}} > 0 \qquad \qquad \square$$

Hypothesis testing I use Remark 1 to construct a test for stability (time-invariance) of consumer beliefs.

H_0 *(Stable consumer beliefs)*: For any given inexperienced consumer, the correlation between her product choices and the choices made by a separate group of already experienced consumers stays the same across the purchase occasions. i.e.,

$$\frac{\partial \ corr(\boldsymbol{s}_i, \boldsymbol{s}_{\mathcal{E}})}{\partial n_i} = 0 \quad \forall i \in \{i | n_i < \overline{n} \text{ and } i \notin \mathcal{E}\}.$$

H_1 *(Unstable consumer beliefs)*: For any given inexperienced consumer, the correlation between her product choices and the choices made by a separate group of already experienced consumers changes as she accumulates more purchase experiences. i.e.,

$$\frac{\partial \ corr(\boldsymbol{s}_i, \boldsymbol{s}_{\mathcal{E}})}{\partial n_i} \neq 0 \quad \forall i \in \{i | n < \overline{n} \text{ and } i \notin \mathcal{E}\}.$$

To test the hypotheses, I use the following step:

1. I create $\mathcal{E} = \{$a sample of 5000 experienced customers who have made at least 15 past visits$\}$.[1] Similarly, I create $\mathcal{I}_{\overline{n}} = \{$a sample of customers who make total $\overline{n}$ of visits in

---

1. The sample size is approximately 25% of the consumers with more than 15 visits in the data set. I

the data period}. I try two different $\bar{n}$ to capture potentially two very different sets of customers: $\bar{n} \in \{2$ (those customers who only visited 2 times during the data period), $10+$ (those customers who visited more than 10 times during the data period)}. I construct $\mathcal{E}$ and $\mathcal{I}_{\bar{n}}$ such that the two sets are mutually exclusive, i.e., $\mathcal{E} \cap \mathcal{I}_{\bar{n}} = \emptyset$. In other words, any transactions made by members $\mathcal{E}$ prior to 15th visits are taken out from the data for the analysis.

2. For each concert, construct $s_{\mathcal{E}j}$ that represents experienced consumers' choice share of a given concert. $s_{\mathcal{E}j}$ for concert j is calculated as

$$s_{\mathcal{E}j} = \frac{\sum_{i \in \mathcal{E}} y_{ij}}{\sum_{i \in \mathcal{E}} \sum_{k \in \{k | Year_k = Year_j\}} y_{ik}} = \frac{\text{\# concert } j \text{ ticket purchases made by } \mathcal{E}}{\text{Total \# ticket purchases made by } \mathcal{E} \text{ in Year}_j}$$

where $y_{ij} = 1$ if consumer i purchases tickets for concert j.

3. Run the following regression with individual fixed effects (α_i) using a sample group of inexperienced consumers $\mathcal{I}_{\bar{n}}$:

$$(s_{\mathcal{E}j} \text{ of concert } j \text{ chosen by consumer } i \in \mathcal{I}_{\bar{n}} \text{ at visit } n < \bar{n}) = \alpha_i + \beta_n n + \eta_{ijt} \quad \forall i \in \mathcal{I}.$$

$$(4.1)$$

4. Check whether β_v is statistically different from 0.

$$H_0 : \beta_n = 0 \quad \text{vs.} \quad H_1 : \beta_n \neq 0$$

Test results Figure 4.1 shows the evolution of correlation between the choices by already experienced customers and by inexperienced customers over time (i.e., $corr(s_{\mathcal{E}}, s_{\mathcal{I}_{\bar{n}}})$ in Remark 1 and 2). Each line denotes the correlation between the vector of concert choices by already experienced customers ($s_{\mathcal{E}}$) and that by inexperienced customers with $\bar{n}$ total visits

---

try different thresholds ranging from 15 to 30, and the correlation between the measures of experienced consumers' choices with different thresholds is greater than 0.93.
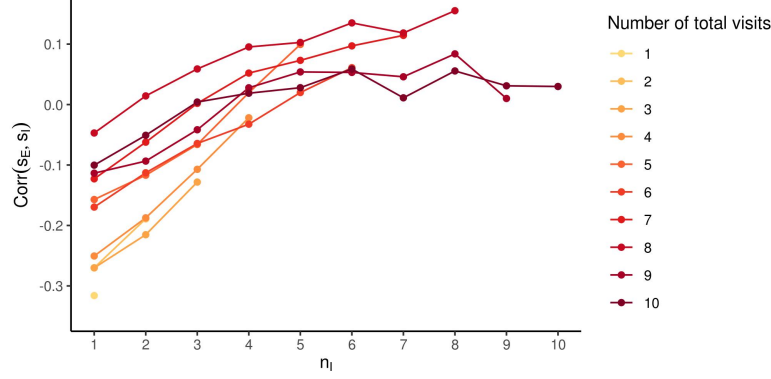
Figure 4.1: Evolution of correlation between the experienced and inexperienced consumers' choices over time

during the data period $(s_{\mathcal{I}_{\bar{n}}})$. For each line, customer base stays the same across the number of visits (x-axis) with no customer entry or exit. Upward-sloping patterns for all customer groups with different total visits imply that, as inexperienced customers make more visits, their concert choices look more like the choices of already existing experienced customers.

Table 4.1 reports the test results. I run a regression specified in Equation (4.1) separately on the choices of consumers with different total visits to see whether the choice resemblance emerges only in a specific group of consumers. The results with statistically significant β_n's in all cases reject the null hypothesis that the correlation between the experienced and inexperienced consumers' choices stays constant across inexperienced consumers' purchase occasions.

4.2   Interpretation of the test result

The test result provides evidence that consumer preferences or beliefs about product values are not stable over purchase occasions. There can be various factors that drive such pattern. Consumer learning about product values rationalizes the pattern of changing beliefs. Targeted advertisement that promote different products to customers based on their length of consumption history can also justify the same pattern. In the specific context of this paper, the targeted advertisement effect is ruled out given that most marketing materials provided

25

Table 4.1: A test for stability of consumer beliefs: Results

	Dependent variable:	
	$s_{\mathcal{E}j}$ chosen at each visit (normalized; divided by mean $s_{\mathcal{E}j}$)	
	(1) Consumers with 2 visits	(2) Consumers with ≥ 10 visits
β_n (Number of visits)	0.017***	0.011***
	(0.004)	(0.001)
Observations	83,122	45,332
R^2	0.730	0.490

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

by the symphony center do not highlight different concerts based on consumer loyalty.

Although the test does not accept consumer learning, it rejects that the data can be explained solely by the model of fixed preferences under perfect information (Remark 1). In other words, what the test result rules out is not that time-invariant heterogeneous preferences do not exist at all in the data; instead, it rules out the non-existence of time-varying consumer beliefs. Figure 4.1 delivers the same conclusion; the increasing resemblance between the choices of inexperienced and already experienced customers implies that the observed consumption pattern cannot be fully explained only by time-fixed heterogeneous preferences. Moreover, the pattern is consistent with what standard consumer learning models predict (Remark 2). This framework of hypothesis testing can be useful in future empirical research at checking whether the data is suitable for studying consumer learning.

Based on the implication from the test, the next chapter shows that the underlying values of available concerts can be identified from experienced consumers' choices under the standard model of consumer learning.

CHAPTER 5

IDENTIFICATION OF CONCERT VALUES

Given the test result that is consistent with the prediction under standard learning models, I show that the true product values, or concert values in this paper's context, can be inverted from the experienced consumers' choices under the widely used assumption in the empirical learning literature. After proposing an estimator of the true concert values based on the invertibility condition ([8]), I discuss the advantage of estimating product values using a subgroup of customers and new insights that this approach can bring to the empirical learning literature.

5.1 An estimator of product values

Given the test results that reject the model of fixed preferences under perfect information, I identify underlying product (concert) values using the assumption of standard learning models (A3).

Using *A1-A3*, I show that the true product values can be inverted from the observed product choices of experienced consumers (Remark 3).

Remark 3. (Invertibility) Suppose A1-A3 hold. Then, there is a one-to-one mapping between the true product values ($\boldsymbol{Q}^*$) and the observed product choices of consumers who have sufficiently high number of past consumption experiences ($\boldsymbol{s}_{\mathcal{E}}$).

Sketch of proof: Let $\boldsymbol{Q}_{in}$ denote a vector of consumer i's beliefs about product values after n visits (purchase occasions). A3 implies that there is $\overline{n}_\delta$ s.t. $|\boldsymbol{Q}_{i\overline{n}_\delta} - \boldsymbol{Q}^*| < \delta \quad \forall \delta > 0$. For each $\delta > 0$, let $\mathcal{E}_\delta$ denote a set of experienced consumers $\{i|n_i \geq \overline{n}_\delta\} = \{i||\boldsymbol{Q}_i - \boldsymbol{Q}^*| < \delta\}$. Then, $\forall \delta \, \exists \sigma_\delta$ s.t. $|\mathscr{S}(\alpha_{\mathcal{E}_\delta}, \boldsymbol{Q}_{\mathcal{E}_\delta}) - \mathscr{S}(\alpha_{\mathcal{E}_\delta}, \boldsymbol{Q}^*)| < \sigma_\delta$ and $\sigma_\delta \to 0$ as $\delta \to 0$. The invertibility of the product share function $\mathscr{S}(\alpha, \boldsymbol{Q}_{\mathcal{E}}) \approx \mathscr{S}(\alpha, \boldsymbol{Q}^*)$ to recover $\boldsymbol{Q}^*$ follows from the proof

by [8].

Using the invertibility condition, I propose an estimator of underlying product values as a function of experienced consumers' choices.

An estimator of concert values Suppose that 1) the random utility component follows i.i.d. Type 1 Extreme Value distribution, 2) consumers make a choice of purchase-no purchase for each concert,[1] and 3) there is no random coefficients within the group of experienced consumers ($\mathcal{E}$). Then, $\widehat{Q}_j^*$ is an estimator of the underlying product value Q_j^* (Berry 1994):

$$\widehat{Q}_j^* = \ln(s_{\mathcal{E}j}) - \ln(s_{\mathcal{E}j0}) - (\text{Hour, Day, Month, Genre Fixed Effects})$$
$$\text{where} \quad s_{\mathcal{E}j} = \frac{\sum_{i \in \mathcal{E}} y_{ij}}{\sum_{i \in \mathcal{E}} \sum_{k \in \{k | Year_k = Year_j\}} y_{ik}} \quad \text{and} \quad s_{\mathcal{E}j0} = 1 - s_{\mathcal{E}j}$$
$$\text{and} \quad \mathcal{E} = \{i | n \geq \overline{n}\}. \tag{5.1}$$

This estimator, computed with the assumption of no heterogeneity, identifies the average preference of each product perceived by the market. More specifically, it reflects the ranking of each concert in a given year in a given genre, voted by a holdout group of already experienced customers, net of time effects that may affect general visitor traffic. The estimator can be easily modified to explicitly account for heterogeneity in perceived product values. One possible approach - clustering experienced consumers using their purchase sequences and constructing a separate measure within each cluster - is introduced in Section A.3 in the Appendix. For subsequent analyses, I use $\widehat{Q}_j^*$ estimated without consumer heterogeneity (following Equation (5.1)) as a proxy for market average perceived value of concert j.

The distribution of the estimated concert values using 5000 experienced consumers with

---

1. This assumption is based on the data pattern that about 10% of experienced consumers consume multiple concerts in a given week. Changing the decision setting to be weekly does not change the estimated measures much.
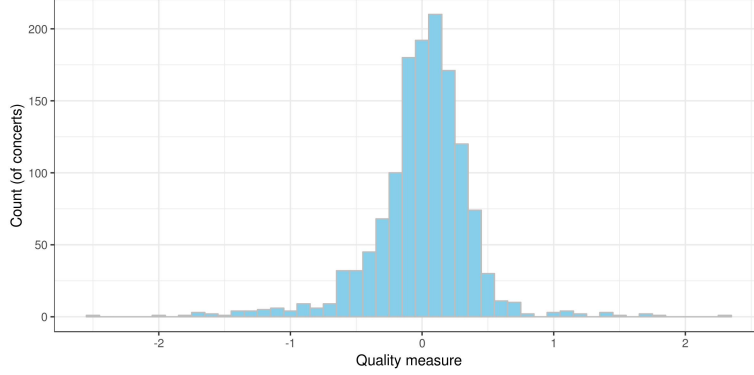
Figure 5.1: Histogram of estimated concert values

$\overline{n} = 15$ is summarized in Figure 5.1.[2] Section A.1 further discusses the specification of the estimator and its implication. Section A.2 discusses the validity of the estimated measure, in which I compare the estimated concert values with billboard rankings of featured artists and composers as well as with average waiting time for each concert.[3]

5.2   Discussion: Benefits of identifying product values

To my knowledge, it is new to the field to identify unobservable product qualities or match values by applying the prediction of standard learning models to the choices of experienced consumers. However, the idea of using choices by individuals outside the group of interest has been applied in prior research to control for unobservables. For instance, [44] use the average national weekly box office sales to control for the qualities of movies screened in a local theater. When estimating the effect of public school quality on residential choices, [15] controls for neighborhood unobservables by using the residential choices of people without school-age children. [1] and [55] use purchases of a certain group of consumers as a predictor of new product failures in various categories.

Constructing a proxy for product values using revealed preferences of informed consumers

---

2. As in Chapter 4.1, I use different values of $\overline{n}$ from 15 to 30 for robustness check, and different cutoff values give highly correlated estimates (correlation greater than 0.93).

3. Waiting time means the number of days between the date of ticket order and the date of performance.

can be useful for several reasons. First, it is hard to find survey data on product qualities that can be used with matching purchase data. Low availability comes from both lack of survey implementation and low response rate to the implemented surveys. Second, even if available, survey data may not be informative of consumers' true preferences. Prior literature raises concerns for various types of survey bias ([45]), and industries also question the validity of survey results based on their low predictive power ([47]).

Inferring product values from choices of experienced consumers also introduces a new way to study consumer learning. In particular, once product values are estimated, researchers can leverage them to trace consumer learning patterns without assuming any specific learning rule (Figure 5.2). Investigating how fast or slowly consumers' product value choices reach the steady state can inform researchers on how fast consumers learn about the underlying product values.

This approach deviates from how most empirical literature demonstrates consumer learning. Instead of jointly estimating consumer preferences and learning parameters based on specific functional form assumptions (e.g., Bayesian updating rule with normal signals and priors), it first estimates consumers' true preferences using experienced consumers' data and then traces which product values consumers choose over time. Identification of product values relies on the assumption that consumers' beliefs about product values converge to the true product values as they get more experienced, which holds in most standard learning models. However, identification of learning pattern does not require any assumptions on how consumers learn, or more specifically, in which shape or with which updating rule consumer learning takes place. Therefore, this approach can be useful in cases in which learning is likely to take place but not in a way that traditional Bayesian learning models predict. One may raise concerns that there may be potential heterogeneity in preferences either among experienced consumers or between experienced and less experienced consumers. However, the approach can easily incorporate heterogeneity in perceived product values, which can be explored more in future research (as discussed in Section 4.2 and A.3).
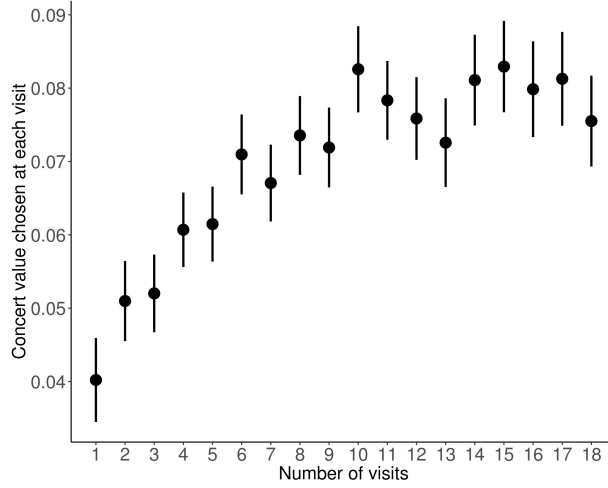
Figure 5.2: Documentation of consumer learning (among consumers who make ≥ 18 visits)

In summary, identifying product values from experienced consumers' choices complements the existing literature by introducing a way to augment unobserved product qualities and to relax certain assumptions about the path of learning.

Next, I run descriptive analyses using the estimated concert values and show how imperfect information about individual concerts causes customer attrition at the symphony center level.

CHAPTER 6

DESCRIPTIVE ANALYSIS

In this chapter, I show descriptive evidence that initial concerts chosen under imperfect information significantly affect subsequent customer attrition at the symphony center level. First, churn rates among first-time visitors vary widely across different concerts, which is consistent with a significant effect of initial concert experiences on churn decisions. Second, I show that the arrival rate of first-time customers is random with respect to the underlying concert values, which suggests incomplete information about the concert values at the early purchase stage. Finally, I show that the concert values experienced in the first few visits are strongly correlated with consumers' subsequent churn decisions even when a rich set of explanatory variables are controlled for.

In the following descriptive analyses, I include only households within 30 miles from the symphony center based on the zip code information.

6.1 Varying retention rates across concerts

Figure 6.1 shows that there is enough variation across concerts in both the number of first-time customers (Figure 6.1(a)) and the rate of subsequent churn among them (Figure 6.1(b)). As Figure 6.1(b) shows, the average retention rate across concerts is about 40%, implying that more than half of the visitors on average do not come back to the symphony center after their first visit. However, the distribution of the retention rates also has a long right tail, meaning that there are a set of concerts after which more first-time visitors stay for subsequent visits.
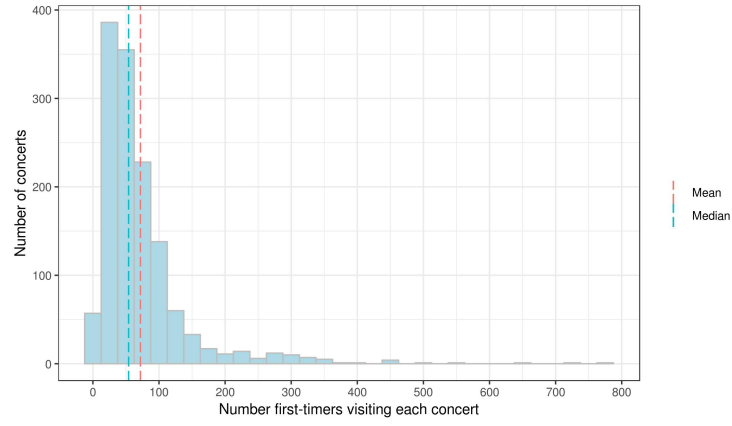
(a) The number of first-timers arriving at each concert
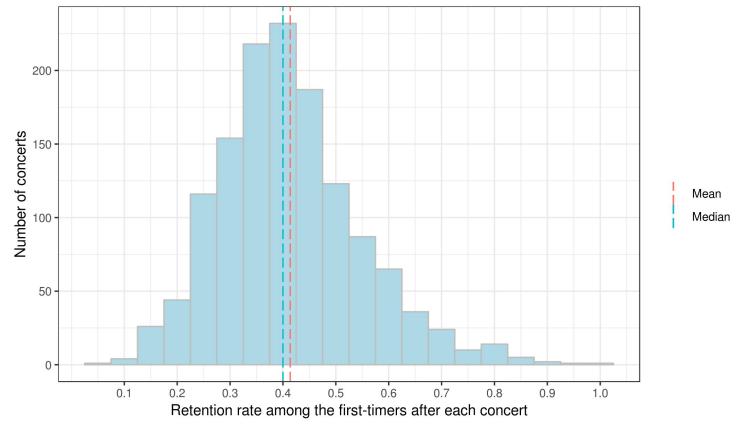
(b) Retention rate among the first-timers after each concert

Figure 6.1: Variations across concerts in the first-time visitors' arrival and retention rates
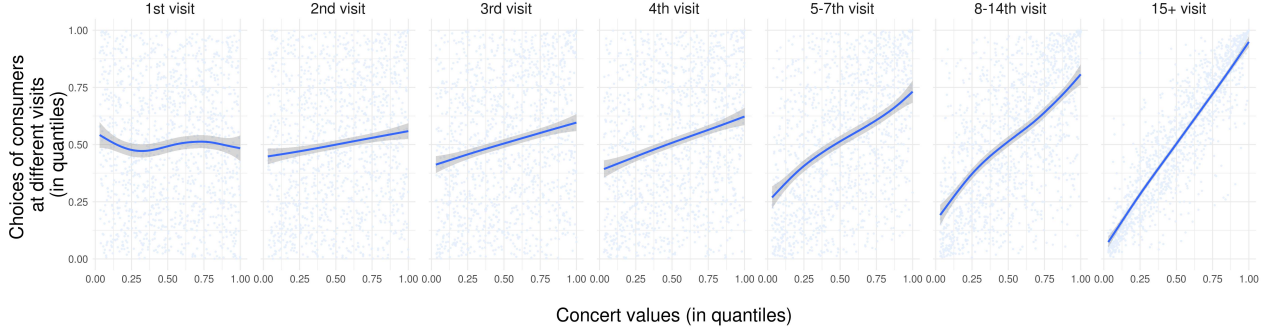
Figure 6.2: Concert values vs. Concert choices by consumers with different numbers of past visits: Quantile-Quantile scatterplots and smoothed conditional means

6.2 Imperfect information about concert values at the initial purchase stage

If the variation across concerts in retention rates described in Figure 6.1 is purely due to customers' fully informed preferences, marketing campaigns like informative advertising about other available concert values may not effectively raise retention. However, if at least some of the observed churn events are attributed to the mismatches between concerts and consumers due to imperfect information, correcting these mismatches by informing customers might favor both visitors and the symphony center.

Table 6.1: Concert choices by consumers with different numbers of past visits: Regression results

	Dependent variable: Concert choice shares (in quantile) among consumers at their						
	1st visit	2nd visit	3rd visit	4th visit	5-7 visits	8-14th visits	15+ visits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Quantile($\widehat{Q}_j^*$)	−0.016	0.132***	0.224***	0.284***	0.470***	0.615***	0.902***
	(0.027)	(0.027)	(0.027)	(0.026)	(0.024)	(0.021)	(0.012)
Constant	0.508***	0.434***	0.388***	0.358***	0.265***	0.193***	0.049***
	(0.016)	(0.016)	(0.015)	(0.015)	(0.014)	(0.012)	(0.007)
Observations	1,350	1,350	1,350	1,350	1,350	1,350	1,350
R^2	0.0003	0.017	0.050	0.081	0.221	0.378	0.814

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

Figure 6.2 shows how customers with different number of past visits choose concerts with respect to the underlying concert values. The x-axis denotes the quantile of the estimated

34

concert values $\widehat{Q}_j^*$, which is created using Equation 5.1 with a holdout sample of 5000 customers with more than 15 visits. The y-axis denotes the quantile of concert choices by customer groups with different levels of experiences, which is constructed using the same Equation 5.1 but with different groups of customers.

Two patterns are noticeable. First, the initial choices by the first-time customers are random with respect to the estimated concert values at the concert level (the first cell of Figure 6.2).[1] Second, customers' concert choices become more positively correlated with the concert values (which is created with a separate sample of already experienced customers) as they visit more, and the correlation becomes positive and significant even on the second visit. Table 6.1 reports the same results with regressions; the correlation between the concert value ranking (quantiles) and the ranking based on the first-time customers' choices is not statistically significant (Column (1)), and the correlation increases as visitors become more experienced.

One can argue that the pattern of choice convergence towards high-value concerts in Figure 6.2 or Table 6.1 is solely due to the systematic change in customer base; consumers with low customer lifetime value might have preference for specific types of concerts that regular consumers do not seek for, and they drop out from the customer base after a few visits, which may create a similar choice resemblance pattern. However, I show in Section 4 that the same choice convergence pattern is observed over visits even when customer base is fixed. This undermines the plausibility that the randomness of concert choices by the first-time visitors, followed by choice resemblance starting from the second visit, is created solely due to systematic differences in consumer preferences.

In summary, there exists incomplete information about true concert values at the ticket purchase stage among inexperienced consumers, which is demonstrated via lack of correlation between the first-time visitors' choices and the estimated concert values. Next, I analyze whether the experienced concert values chosen under imperfect information has any impact

---

1. Result remains the same when I use raw scales instead of quantiles.

on subsequent churn at the symphony level.

6.3 The effect of experienced concert value on customer churn

Given the suggestive evidence that first-time visitors do not have complete information about underlying concert values when they purchase tickets, I look at whether the experienced concert values chosen under imperfect information affect the customers' subsequent churn at the symphony center level.

First, to analyze the effect of experienced concert values on following churn, I run a regression with individual-level churn decision as an outcome variable and the recent concert value experienced as the key explanatory variable:[2]

$$Pr(Churn_{i\tau} = 1 | \tau\text{-th visit}) = \frac{\exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)}{1 + \exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)} \quad (6.1)$$

where $\widehat{Q}_{i\tau}$ is the estimated concert value that i experienced at her τ-th visit. Similarly, $\widehat{Q}_{i\tau-1}$ and $\widehat{Q}_{i\tau-2}$ are the lag concert values experienced at $\tau - 1$-th and $\tau - 2$-th visit respectively. To ease the interpretation, I normalize $\widehat{Q}_{i\tau}$ to have mean 0 and standard deviation of 1. Table 6.2 describes a set of control variables ($Z_{i\tau}$) included in the regression.

$$Z_{i\tau} = \{ \log(\text{Seat Quantity}_{i\tau}), \log(\text{\# Days purchased in advance}_{i\tau}), \log(\text{Price paid}_{i\tau}),$$

$$\text{Concert hour}_{i\tau}, \text{Concert day of week}_{i\tau}, \text{Concert month}_{i\tau}, \text{Concert genre}_{i\tau},$$

$$\text{Popularity of the concert among the first-time customers}_{i\tau},$$

$$\text{Zip code}_i\} \quad (6.2)$$

---

2. In the reported results, churn is defined to be 1 if a consumer does not come back to the symphony center at all within the data period. Given the burn-in and burn-out periods I use, churn = 1 means not coming back to the symphony center at least in four years. Several other specifications of churn (e.g., churn defined as not coming back for n year with $n \in \{1, 2, 3, 4\}$) give qualitatively the same results.

Table 6.2: List of control variables

Variable	Unit	Type	# Levels (if any)	Description
Seat Quantity	Consumer-Concert	Discrete	-	Number of seats purchased for a given concert
# Days purchased in advance	Consumer-Concert	Discrete	-	Number of days between performance date and ticket purchase date
Price paid	Consumer-Concert	Continuous	-	Ticket price paid (in dollars) net of discounts/ promotions
Concert hour	Concert	Categorical	3	Performance start time (Morning (- 12 PM); Afternoon (12 PM - 6 PM); Evening (6 PM -))
Concert day	Concert	Categorical	7	Day of week (Sunday to Monday)
Concert month	Concert	Categorical	12	Month (January to December)
Concert genre	Concert	Categorical	15	Casual classic; Casual fusion; Emerging professionals; Emerging professionals, fusion; Chamber; Main (orchestra); Family; Jazz; Movies; Non-western; Guest chamber; Guest contemporary; Guest orchestra; Guest piano; Specials
Popularity among the first-timers	Concert	Continuous	-	The measure is created using Equation (5.1) but using the choices of the first-time customers instead of the choices of sample experienced customers.
Zip code	Consumer	Categorical	256	Any zip code that contains less than 200 customers are aggregated into "Others". (21% of the customers)

Detailed data in each transaction allows me to control for various alternative explanations that may cause spurious correlation between the experienced concert values and subsequent churn decisions. For example, including the price paid in each ticket order controls for potential indirect effect of price on churn decision via concert value. The number of days between the date of ticket order and the date of performance absorbs the effect of ticket availability at the moment of purchase. Concert genres control for potential differences in the underlying churn rates across customers who prefer different genres (e.g., the difference in churn rates between movie sound tracks concert visitors and chamber music concert visitors).

One might expect that there are certain concerts that are popular among one-time customers, and that choosing these concerts is positively correlated with subsequent churn and is negatively correlated with regular consumers' tastes. If such concerts exist and are not controlled for, negative correlation between concert values and following churn may not reflect a causal impact of experienced concert values on churn but a mere correlation between the two. To control for this mechanism, I include a measure of popularity among the first-time customers (i.e., controlling for those concerts that are only tailored for casual visitors). This measure of popularity is created using Equation (5.1) but with the choices of first-time customers (instead of the choices of sample experienced customers). If there is no causal impact of experienced concert value on churn, then the coefficient of recent concert value (β_{q1} in Equation 6.1) would become insignificant after controlling for all these confounders.

Table 6.3 summarizes the results. Significant and negative coefficients of concert values, even when a list of confounding factors are controlled for, imply that there is a systematic relationship between the experienced concert values and the subsequent churn rates: the higher the experienced concert value is, the lower the chance that a consumer churns after the visit. Exclusion of control variables do not create any qualitative difference in the regression results as Column 4-6 show, indicating that the effect is not driven by changes in the composition of consumer demographics.

A statistically significant coefficient of the most recent concert value when the lag concert

Table 6.3: The effect of consumption experience on subsequent churn decision (Binary logit model)

	Dependent variable:					
	Churn = 1 after τth visit					
	(1) After 1st visit	(2) After 2nd visit	(3) After 3rd visit	(4) After 1st visit	(5) After 2nd visit	(6) After 3rd visit
$\widehat{Q}_{i\tau}$ (normalized)	−0.071***	−0.064***	−0.037*	−0.116***	−0.142***	−0.106***
	(0.006)	(0.012)	(0.019)	(0.005)	(0.018)	(0.019)
$\widehat{Q}_{i\tau-1}$ (normalized)		−0.034***	−0.040**		−0.113***	−0.111***
		(0.012)	(0.020)		(0.011)	(0.019)
$\widehat{Q}_{i\tau-2}$ (normalized)			−0.007			−0.077***
			(0.020)			(0.018)
Popularity among the first-time visitors	0.046***	0.041***	−0.012			
	(0.007)	(0.014)	(0.020)			
log(Seat Quantity)	0.238***	0.161***	0.257***			
	(0.019)	(0.035)	(0.053)			
log(# Days purchased in advance)	−0.121***	−0.142***	−0.175***			
	(0.005)	(0.010)	(0.014)			
log(Price paid)	−0.018***	−0.027**	0.011			
	(0.006)	(0.012)	(0.018)			
Observations	96,370	30,252	17,555	96,370	30,252	17,555
Log-likelihood	-59404	-16494	-8072	-63696	-18670	-8948
Controls ($Z_{i\tau}$)	Y	Y	Y	N	N	N

Note: *p<0.1; **p<0.05; ***p<0.01. Controls include zip code, time, and genre fixed effects.

39

values are also included further supports the causal impact of experienced concert value on churn. If we assume that the first concert value chosen by any consumer informs her latent type (e.g., long tenure vs. short tenure) and her churn decision is solely driven by her latent type, controlling for the concert value chosen at the first visit should remove the effect of the most recent concert value on the churn decision after her second or third visit. Under this assumption, the persistent effect of the most recent concert value on customer churn even after controlling for the previous concert choices implies the causal impact of experienced concert value on consumer decisions to churn.

As a robustness check, in Section A.4, I run a linear probability model using the same outcome variable and a richer set of control variables. In addition to the control variables listed above in Equation 6.2, the following variables are included in the regression: seating area at the zone level (e.g., box seats, main floor, upper balcony), type of price promotion (e.g., student discounts, web 15% coupon), and ticket sales channel (e.g., box office, website). In this specification, any differential impact of seat locations, consumer demographics, and other concert visit behavior on customer retention is further absorbed. Figure A.4.3 in Section A.4 shows that there still exists a significant negative correlation between the experienced concert value and the subsequent churn rate.

The results imply that initial concert experiences can create an economically significant impact on customer retention. The coefficient in Column 1 implies that, given that the average churn rate after the first visit is 61.3%, 1-standard-deviation increase in the first experienced concert value leads to the reduction of churn rate from 61.3% to 59.5%. Figure 6.3 plots the distribution of predicted increase in retention rate if every first-time customer was treated with the highest-value concerts. It shows that the retention rate can go up by 10 percentage points on average under this treatment.

To get a sense of whether a decrease in churn rate creates any meaningful impact on the symphony center's revenue, I approximate the symphony center's ticket revenue under different churn rates using the previous regression results (Table 6.4). For each consumer

40
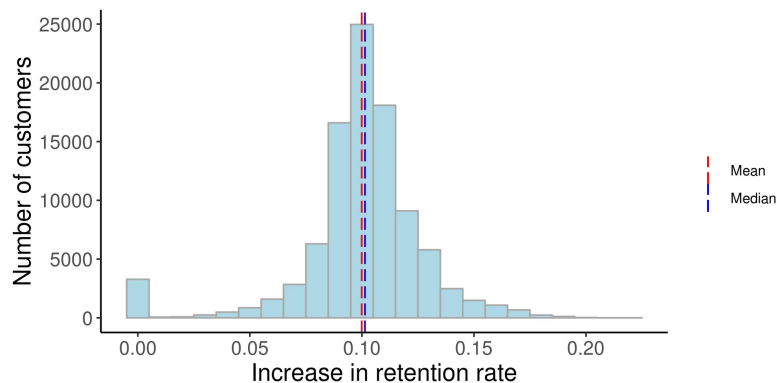
Figure 6.3: Predicted increase in retention rate after the first visit if customers visited the highest-value concerts on their initial visits

Table 6.4: Predicted increase in ticket revenue if customers visited the highest-value concerts on their initial visits
(in % of the ticket revenue from the first visit)

	2009	2010	2011	2012	2013	2014	2015	Total
2009 cohort	4.53%	5.18%	4.02%	4.05%	3.63%	3.17%	3.10%	27.66%
2010 cohort	-	4.06%	4.29%	3.84%	3.45%	3.30%	3.03%	21.98%
2011 cohort	-	-	3.32%	4.21%	3.63%	3.02%	3.04%	17.21%
2012 cohort	-	-	-	2.86%	3.72%	3.26%	2.80%	12.63%
2013 cohort	-	-	-	-	2.48%	3.37%	2.99%	8.84%
2014 cohort	-	-	-	-	-	2.23%	3.31%	5.55%
2015 cohort	-	-	-	-	-	-	4.10%	4.10%

cohort (defined by which year they make first visit), I calculate how ticket revenue changes if the highest-value concert was offered in their first visits, assuming that the churn rate at each visit beyond the first visit stays intact.[3] On average, each cohort treated with high-value initial concerts creates about additional 3% of the first-visit ticket revenue every year. This highlights that curating initial concert experiences can be a one-time investment that can generate a flow of revenue onward.
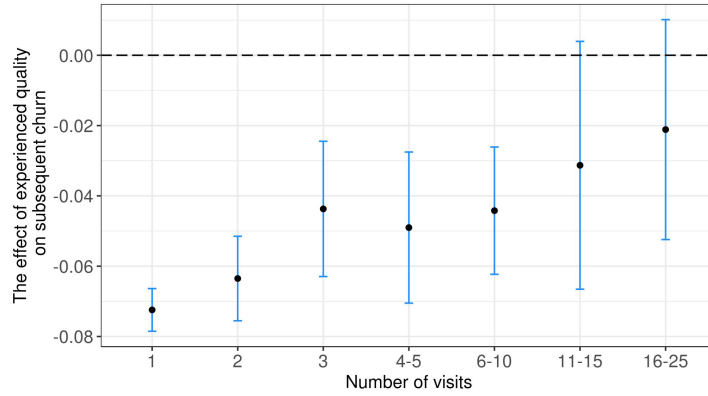
Figure 6.4: The effect of recent concert value on subsequent churn over visits

Figure 6.4 shows how the effect of recent concert value on subsequent churn decision changes over visits. The x-axis denotes the number of visits, and the y-axis denotes the effect of most recently experienced concert value $(\widehat{Q}_{i\tau})$.[4] As implied by a standard learning model, the effect of experienced concert value diminishes and becomes statistically insignificant as the number of visits goes up. It emphasizes that the few initial concert experiences have larger impact on immediate customer churn than any other later visits do.

Initial concert experiences also affect other factors that determine customer lifetime value, including the length of tenure or the amount of donation. In Section A.5, I run the same regression as in Equation (6.1) using two alternative outcome variables: conversion to a

---

3. The table is calculated using the following equation:

$$\Delta\text{Ticket Revenue in Year } t \text{ from consumer cohort } c = \text{Median}\Delta\text{Churn rate reported in Figure 6.3}$$
$$\times (\text{Observed ticket revenue generated by } c \text{ in Year } t).$$

4. I include up to 3 lag concert value variables in this specification.

regular consumer (with 10+ visits) and conversion to a donor.[5] The results imply that the initial concert experience is correlated with both the probability of being a regular customer beyond the second visit as well as customers' subsequent donation behavior. In other words, the overall effect of initial concert experiences on customer lifetime value would be large enough to make significant contribution to the symphony center's profit.

In summary, the causal impact of experienced concert value on customer retention is supported by two joint pieces of evidence: 1) the randomness of the initial concert choices made by new consumers (Figure 6.2 and Table 6.1), and 2) the correlation between the experienced concert values and customer churn in Table 6.3, which stays significant even when a rich set of potential confounders are controlled for. The effect of initial concert experience on customer lifetime value is even greater when its impact on customer tenure and donation is also taken into account.

6.4   Discussion

The findings suggest that customers who are exposed to low-value concerts (due to imperfect information at the purchase stage) simply abandon the symphony center afterwards instead of trying other concerts. This pattern may arise because they generalize the experience to the center's other concerts they have not tried. In other words, two layers of imperfect information - one about the underlying concert values at the purchase stage and the other about the range of available concert values at the updating stage - can jointly lead to a significant impact of initial concert experience on customer retention at the symphony center level.

Another important pattern in the data is that, even when new concert features are introduced, visitors become better at choosing high-value concerts over time (as implied in Section 4.1) although they have not experienced those concert features before. This implies

---

5. I use a data set on donation activities that shares a unified customer ID system with the ticket purchase data set.

that consumers collect information about the values of those new features not through direct consumption but through a separate information channel.

Despite the richness of the existing learning models, most traditional Bayesian learning models cannot fully rationalize these data patterns or, if they can, are computationally intractable to estimate. To offer clear motivations for the new framework, I discuss in Section A.6 in the Appendix whether a traditional Bayesian learning framework can explain the descriptive patterns and what its potential limitations are. The main points can be summarized as follows: 1) a Bayesian learning model with forward-looking behavior would not rationalize consumers' abandonment of the symphony center after a single visit because any rational consumer would need at least two samples to resolve the uncertainty about the distribution of concert values (unless the trial cost is too high), 2) if consumers are myopic, a firm (symphony center)-level learning model should allow for a very high signal variance to justify the high churn rate after a single visit (instead of normalizing it to 1 as in many previous empirical studies), 3) a product-level learning model that can incorporate most of the findings is computationally burdensome given the large product and feature space, and 4) even the most-detailed attribute-level learning cannot fully explain how consumers can become better at choosing high concert values over visits even among the concerts with new (untried) features.

The next chapter proposes a framework of consumer learning with two information acquisition processes that rationalize the descriptive findings.

CHAPTER 7

A FRAMEWORK OF CONSUMER LEARNING

I propose a new structural model of consumer learning that incorporates two types of incomplete information, one at the purchase stage and the other at the updating stage. The framework also allows for two information acquisition activities discussed in Section 3: search and purchase. Experiences from purchases lead to learning spillovers and affect subsequent search decisions. Search activities allow consumers to become better at selecting a high-value product even when they are faced with new features they have not experienced before.

The model allows for flexible patterns of learning spillovers in a computationally tractable way. Consumers extrapolate their past experiences to other untried products by taking the weighted average of the past experiences, in which the weight is assigned based on the product similarity between the previously experienced products and the untried products of interest.

(a) Signal from the first purchase (b) Signal from the second purchase (c) Resulting belief after the two purchases
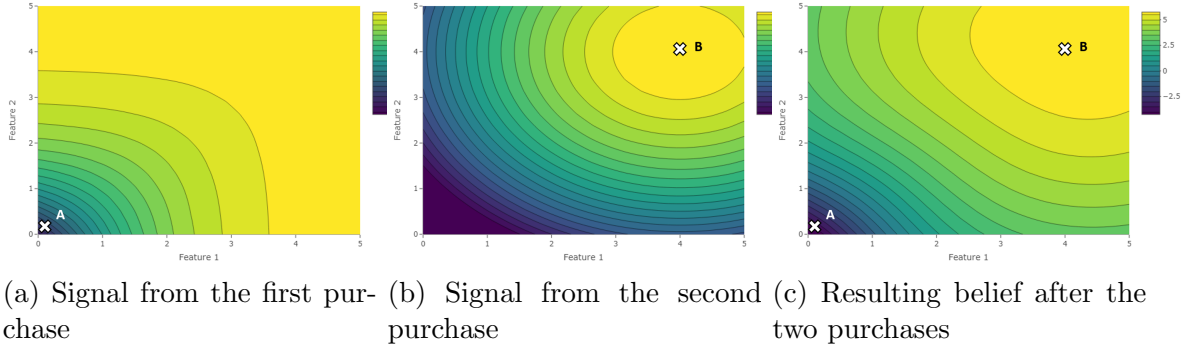
Figure 7.1: Illustration of learning spillover from consumption

Figure 7.1 shows an illustrative example of how learning spillovers from consumption experience take place in a product feature space. Two axes denote two available product features, and different colors denote the range of predicted consumption values for each combination of the two product features. The brighter the color is, the higher the predicted consumption value is. Initially, a consumer tries product A and experiences a low value from it. Given the experience, she updates her prediction on all the other combinations of product

features based on their distances from A in the feature space (Figure 7.1(a)). In the next purchase, she tries product B which is predicted to have a high value based on her previous experience. Suppose that product B indeed gives high-value experience (Figure 7.1(b)). The resulting belief about all the other untried products after the two purchases is the weighted average of the predicted values in (a) and (b) (Figure 7.1(c)).

To rationalize the data pattern that consumers become better at selecting high-value products even when they face new product features they have not tried before, I also allow consumers to obtain information through an additional channel besides purchase, which I call search. Again, search in this framework means any information acquisition behavior not limited to online search, ranging from reading pamphlets about upcoming concerts to searching for online reviews of guest performers in this empirical context. Decision to search is determined by how pleasant the previous consumption experiences are; the better the prior experiences are, the higher the incentive for search is in the next period. This creates an incremental impact of prior experiences on subsequent purchase behavior, since the amount of information that customers obtain before making purchase is a function of past consumption experiences.

Consumers receive signals from both consumption and search activities. Based on the acquired information from both channels, consumers form predicted consumption utility and make a choice that maximizes their predicted utility. I assume that consumers are myopic, i.e., they maximize the current-period predicted consumption utility. [1]

7.1   Utility

Consumer i's utility (u_{ijt}) from consuming product j in time period t is given by

---

[1]. Each concert visit requires relatively high trial cost including time and monetary expense, which makes the assumption of myopic learning more plausible.

$$u_{ijt} = U(Q_j^*, \epsilon_{ijt})$$

$$u_{i0t} = U(u_0, \epsilon_{i0t}) \tag{7.1}$$

where Q_j^* is the true value of product j and ϵ_{ijt} is the random component of consumption utility that is known to i but is unobservable to researchers. u_{i0t} refers to the utility from consuming an outside option.

Given that there is imperfect information on the product values of interest at the purchase stage, consumer i makes purchase decisions based on predicted consumption utility:[2]

$$\widetilde{u}_{ijt} = U(\widetilde{Q}_{it}(X_j), \epsilon_{ijt})$$

$$\widetilde{u}_{i0t} = U(u_0, \epsilon_{i0t}) \tag{7.2}$$

where $\widetilde{u}$ denotes predicted consumption utility and $\widetilde{Q}_{it}(X_j)$ denotes product value predicted by consumer i at time t given the observable product characteristics X_j.

7.2   Updating beliefs

7.2.1   Updating beliefs about product values through consumption

After consuming product j, consumer i uses the difference between her prediction and the realized true value to update her belief about all the untried products. Let $B_{it}(X_j)$ denote the prior belief of product j's value given the observable product features X_j, which is carried over from the previous consumption experiences. Q_j^* refers to the true value of product j that is realized after consumption. Let Δ_{ijt} denote the discrepancy between the true value

---

2. I refrain from using the term 'expectation' because here consumer belief is not modeled as a distribution.

and her belief of the product value:

$$\Delta_{ijt} = Q_j^* - B_{it}(X_j). \tag{7.3}$$

Then, consumer i's belief of the value of an untried product k is updated based on the realized discrepancy (Δ_{ijt}) and the similarity between j and k defined in product feature space (denoted by $d_{jk} = d(X_j, X_k)$):

$$B_{it+1}(X_k) = B_{it}(X_k) + g(N_{it}, d_{jk})\Delta_{ijt} \tag{7.4}$$

$$= B_{it}(X_k) + g_{ijkt}(Q_j^* - B_{it}(X_j)). \tag{7.5}$$

Here, similarity between product j and k is known to both consumers and researchers based on the observed product features X_j and X_k.[3] $g_{ijkt} = g(N_{it}, d_{jk})$ determines how much consumer i generalizes the information from a single experience (Δ_{ijt}) to other untried products based on product similarities (d_{jk}). If $\frac{\partial g}{\partial d} = 0$, then the discrepancy Δ_{ijt} is equally applied to the beliefs of all the untried products regardless of how different the products are from the previously consumed product (j in this case). If $\frac{\partial g}{\partial d} < 0$, learning spillover from the consumed product j is stronger for those products that are more similar to j. I assume that $|g(\cdot)| \leq 1$, i.e., the weight on the information spillover from consumption cannot exceed 1.

7.2.2 Updating beliefs about product values through search

Given the updated beliefs from consumption (B_{it}), consumers can adjust their predictions on product values via search.

---

3. The model does not allow heterogeneity across consumers in perceived similarities between products.

$$\widetilde{Q}_{it}(X_k) = B_{it}(X_k) + S_{it}(X_k)$$

$$= B_{it}(X_k) + \phi_{ikt}(Q_k^* - B_{it}(X_k)) \tag{7.6}$$

$$= \phi_{it}Q_k^* + (1 - \phi_{it})B_{it}(X_k) \tag{7.7}$$

where ϕ_{it} = information gain through search ("search intensity")

$$= \phi(\text{expected consumption benefit, expected search benefit,} \tag{7.8}$$

$$\text{the amount of information the firm sends to } i). \tag{7.9}$$

Search intensity is a function of i's motivation for search, which is endogenously determined by expected consumption benefit and search benefit. It is also affected by the amount of information that a firm sends to consumer i via various marketing channels, because search captures all the external sources of product information besides consumption. I assume that, although the amount of information sent out by the firm varies across consumers, the set of products for which information is provided is the same across all individuals. In other words, the same set of products are advertised by the firm to different customers, but the advertising intensity or the amount of details provided for each product can differ across customers.

I assume that $\phi_{it} \in [0, 1]$; $\phi_{it} = 0$ means that there is no further correction of predicted value via search, while $\phi_{it} = 1$ implies that the true value is fully recovered via search. It is $\widetilde{Q}_{it}(\cdot)$, not $B_{it}(\cdot)$, that enters to consumer i's predicted utility when she makes a product choice.

Note from Equation (7.7) that the process is equivalent to acquiring another signal through search and incorporating it into the existing belief in a Bayesian manner. Here, the signal from search is the true value itself (Q_k^*) and the weight attached to the signal is represented by ϕ_{it}, which is endogenously determined by the past consumption experiences. Therefore, the search step described here can also be viewed as a learning process with

heteroskedastic signals whose variance is endogenously determined by previous consumption experiences.

CHAPTER 8

MODEL

Using the framework introduced in Chapter 7, I construct a structural model for the empirical context of the paper.

Each week, consumers choose among the concert offerings of the week and an outside option (i.e., not purchasing any concert ticket). Purchase decision is a function of predicted consumption utility. When they make a purchase and visit a concert, consumption utility is realized; it is a function of an individual's preference over different genres and the true value of the concert visited. After the concert visit, the information she receives from the concert experience is used to update her belief on any upcoming concerts, which is discounted by how different the upcoming concerts are from the one she has visited. The past concert experiences also determine search intensity in the upcoming weeks by forming her expected consumption and search benefit.

8.1 Utility specification

Consumer i's utility (u_{ijt}) from visiting concert j in week t is given by

$$u_{ijt} = U(Q_j^*, x_{jt}, p_{jt}, \epsilon_{ijt}) = \gamma_{ig} + \tau_{ij} + \delta_i \widehat{Q}_j^* + \lambda_i \widehat{F}_j - \alpha_i p_{jt} + \epsilon_{ijt}$$

$$u_{i0t} = U(u_0, \epsilon_{i0t}) = u_0 + \epsilon_{i0t} \tag{8.1}$$

where γ_{ig} is i's preference for genre g that concert j belongs to, τ_{ij} is the month, day, and hour fixed effect, $\widehat{F}_j$ is the popularity of concert j among first-time visitors (constructed in Section 5.3), and p_{jt} is the price of concert j.[1] $\widehat{F}_j$ is included in addition to $\widehat{Q}_j^*$ to capture

---

1. To avoid modeling consumers' seat choices, I use concert-level price (p_{jt}) instead of the actual price paid by individual customers (p_{ijt}) in this model. Concert-level price is the average ticket price for Main Floor seats, which is summarized in Table 3.4.

customers' heterogeneous tastes for specific concerts that are tailored mainly for one-time visitors (e.g., concerts featuring pop music in a botanical garden). δ_i captures i's sensitivity to the estimated true concert value $\widehat{Q}_j^*$. ϵ_{ijt} is the random component of consumption utility that is known to consumer i but is unobservable to researchers. u_{i0t} refers to the utility from consuming an outside option. u_0 is normalized to be 0.

Since there is imperfect information on concert values, consumer i makes purchase decisions based on the predicted consumption utility:

$$\widetilde{u}_{ijt} = \gamma_{ig} + \tau_{ij} + \delta_i \widetilde{Q}_{it}(X_j) + \lambda_i \widehat{F}_j - \alpha_i p_{jt} + \epsilon_{ijt}$$
$$\widetilde{u}_{i0t} = u_0 + \epsilon_{i0t} \tag{8.2}$$

where $\widetilde{u}$ denotes the predicted consumption utility and $\widetilde{Q}_{it}(X_j)$ denotes the concert value predicted by consumer i at week t given the observable product characteristics X_j.[2]

Let $\widetilde{v}_{ijt}$ denote the deterministic component of predicted consumption utility. Assuming that the random utility component ϵ_{ijt} is *i.i.d.* Type 1 Extreme Value, I can write the purchase probability of concert j among the available concerts at week t ($\mathcal{J}_t = \{0, 1, \ldots, J_t\}$) as

$$Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\widetilde{v}_{ijt})}{\sum_{k \in \mathcal{J}_t} \exp(\widetilde{v}_{ikt})}. \tag{8.3}$$

8.2 Specification of belief updating

8.2.1 Updating beliefs about concert values through consumption

After visiting concert j, consumer i uses the difference between the experienced concert value and the predicted concert value to update her belief about the upcoming concerts. Let

---

2. The model assumes that the popularity among first-time visitors, $\widehat{F}_j$, is fully known to consumer i at the purchase stage.

Δ_{ijt} denote the discrepancy between the true and the predicted value based on the previous concert visit experiences (as in Equation 7.3) :

$$\Delta_{ijt} = \widehat{Q}_j^* - B_{it}(X_j).$$

Consumer i's belief on the value of an upcoming concert k is updated based on the discrepancy (Δ_{ijt}) and the similarity between concert j and k (denoted by $d(X_j, X_k)$):

$$B_{it+1}(X_k) = B_{it}(X_k) + g_{ijkt} \cdot \Delta_{ijt} \tag{8.4}$$

$$\text{where } g_{ijkt} = (\exp(\rho_{it} \cdot d(X_j, X_k)))^{-1}. \tag{8.5}$$

g_{ijkt} determines how much consumer i generalizes the information from a single concert experience (Δ_{ijt}); if $g_{ijkt} = 1 \ \forall k$, then the discrepancy Δ_{ijt} is applied to all the upcoming concerts equally regardless of how different those upcoming concerts are from the concert visited. I assume that $g_{ijkt} \in [0, 1]$. I model ρ_{it} to be the following:

$$\rho_{it} = \exp\left(\rho_{i0} + \rho_{i1} N_{it}\right). \tag{8.6}$$

Based on this specification, the scope of learning spillovers from the recent concert experience varies across experience levels. If $\rho_{i1} > 0$, learning spillovers from the previous concert experience take place more locally as consumers make more visits. This means that, as consumers become more experienced, a concert experience only affects consumers' beliefs about the concerts that are similar to the experienced one. If $\rho_{i1} < 0$, spillovers from the recent visit take place more globally as consumers become more experienced. In other words, the previous concert experience affects consumers' beliefs about the upcoming concerts more broadly as consumers become more experienced.

8.2.2 Updating beliefs about concert values through search

Given the updated beliefs carried over from previous concert experiences, consumers adjust their prediction of concert values via search.

$$\widetilde{Q}_{it}(X_k) = B_{it}(X_k) + \phi_{it}(\widehat{Q}_k^* - B_{it}(X_k))$$

where ϕ_{it} = information gain on true concert value via search

$$= \phi(\text{expected consumption benefit, expected gain from search,}$$

$$\text{the amount of concert information the symphony center sends to } i)$$

$$= \phi\left(\frac{1}{N_{it}}\sum_{j\in\mathcal{C}_{it}}\widehat{Q}_j^*, \ \frac{1}{N_{it}}\sum_{j\in\mathcal{C}_{it}}\left[\widehat{Q}_j^* - \frac{1}{N_{it}}\sum_{j\in\mathcal{C}_{it}}\widehat{Q}_j^*\right]^2, N_{it}\right)$$

and $\phi_{it} \in [0,1]$ \hfill (8.7)

$\mathcal{C}_{it}$ refers to the set of concerts that i has visited up to week t. Expected consumption benefit is modeled as the average realized concert value from the previous visits up to week t, and the expected gain from search is modeled as the variance of the experienced concert values up to week t. N_{it}, the number of visits, is included in the search intensity parameter because the frequency consumer i receives informative marketing contents (ranging from direct mails on upcoming concerts to pamphlets they receive when they visit the venue) depends on how many visits she has made so far. $\phi_{it} = 0$ means that there is no further correction in predicted concert values via search, while $\phi_{ijt} = 1$ implies that the true values of concerts available in week t is recovered via active search prior to purchase. I specify ϕ_{it} as follows:

$$\phi_{it} = \frac{\exp(\phi_{i0}a_{it} + \phi_{i1}b_{it} + \phi_{i2}N_{it} + \phi_{i3})}{1 + \exp(\phi_{i0}a_{it} + \phi_{i1}b_{it} + \phi_{i2}N_{it} + \phi_{i3})}$$

$$\text{where } a_{it} = \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} \widehat{Q}_j^*$$

$$\text{and } b_{it} = \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} \left(\widehat{Q}_j^* - \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} \widehat{Q}_j^* \right)^2. \tag{8.8}$$

8.3 Specification of prior knowledge about concert values

To take into account potential differences in prior knowledge about true concert values, I allow that consumers start with different information set which is estimated by the model:

$$\widetilde{Q}_{i0}(X_j) = \phi_{ip}\widehat{Q}_j^*$$

$$\text{where} \quad \phi_{ip} = \frac{\exp(\phi_{i4})}{1 + \exp(\phi_{i4})}. \tag{8.9}$$

If $\phi_{ip} = 1$, it suggests that consumer i starts with perfect information on concert values when they first visit the symphony center. If $\phi_{ip} = 0$, it means that consumer i does not have any information about different values across concerts when making the initial visit.

8.4 Identification

Clear identification strategy is required to decompose learning from search and learning from consumption when only consumption (ticket purchase) data is available. According to the model, consumption experiences do not allow consumers have better information on other high-value concerts that are not particularly similar to the concerts visited. Learning from search, however, allows consumers to have better information about all untried concerts no matter how similar they are to the ones previously visited. Therefore, learning from consumption is identified if consumers are able to select (avoid) high-value (low-value) products

among the concerts similar to the previously consumed, whereas learning from search is identified if consumers are able to select (avoid) high-value (low-value) concerts from an unfamiliar group of concerts.

In particular, two data variations are used for identification: 1) whether or not each consumer returns after a visit, and 2) which product she chooses if she returns. The first variation helps identification of the parameters that govern how consumers generalize the information from a single visit to other untried concerts (ρ_{it}). The second variation can be decomposed into two components. First, which concert value she chooses when she returns determines how much she adjusts her predicted value through active search (ϕ_{it}). More specifically, if a consumer becomes more likely to choose high-value concerts among a set of very different concerts from the previously visited concerts, this pattern identifies how much the consumer updates her value prediction via search. Second, which concert features she chooses when she returns helps identify how local or global the learning spillover from consumption is (ρ_{it}). If she visited a high-value concert and she chooses the next concert to be very similar to the previous one, then the information from the previous visit may have updated her prediction on the upcoming concerts in a local manner, i.e., updating only for the concerts that are similar enough to the previously visited. However, if she chooses a very different concert at her subsequent visit, it implies that the prior pleasant concert experience has globally updated her beliefs about all the untried concerts.

Unlike in many other empirical learning literature, prior knowledge about concert values can be identified because the concert value $\widehat{Q}^*$ is treated as data instead of being jointly estimated in the model. Prior knowledge is identified by rationalizing consumer's first concert choice. For example, if she starts with a low-value concert and chooses higher-value concerts in her subsequent visits, the model will rationalize this pattern by estimating her prior knowledge to be low. On the other hand, if she starts with a high-value concert from the first visit and chooses high-value concerts onward, the model will rationalize this pattern by estimating her prior knowledge to be high.

CHAPTER 9

ESTIMATION

For estimation, I sample 10,000 consumers who are within 30 miles from the symphony center. For each individual, the week in which the first visit is made is set to be $t = 1$ and T is set to be 100 weeks.[1] Product similarity between any given concert pair is computed via weighted Gower distance (See Section A.7 in the Appendix) and is scaled so that the maximum observed product distance is 1.

θ_i denotes a vector of utility parameters that are estimated:

$$\theta_i = \{\{\gamma_{i1}, \ldots, \gamma_{i8}\}, \{\tau_{i,weekend}, \tau_{i,summer}, \tau_{i,evening}\}, \alpha_i, \delta_i, \lambda_i, \{\rho_{i0}, \rho_{i1}\}, \{\phi_{i0}, \ldots, \phi_{i4}\}\}.$$

Heterogeneity in parameters I use demographics and mixtures of normals to model both observed and unobserved heterogeneity in parameters:

$$\theta_i = Z_i \Gamma_{s_i} + \iota_i$$

$$\iota_i \sim N(0, \Sigma_{s_i})$$

$$s_i \sim \text{Multinomial}_K(\pi) \tag{9.1}$$

where Z_i consists of household income and travel distance inferred by zip code and K denotes the number of normal mixtures. I estimate both $K = 2$ and $K = 3$.

Priors are defined as follows:

$$\Gamma \sim N(\overline{\Gamma}_{s_i}, A_{\Gamma_{s_i}}^{-1})$$

$$\pi \sim \text{Dirichlet}(\overline{\alpha})$$

$$\Sigma_s \sim IW(\nu, V). \tag{9.2}$$

---

1. The data shows that the churn rate stabilizes approximately 100 weeks after a consumer cohort enters.

To make draws from the posterior of $\theta \sim N(Z_i \Gamma_{s_i}, \Sigma_{s_i})$, I define an MCMC chain to be the following:

$$\theta_i | s_i, Z_i \Gamma_{s_i}, \Sigma_{s_i} \tag{9.3}$$

$$\pi, s, \{\Gamma_s\}, \{\Sigma_s\} | \{\theta\}. \tag{9.4}$$

I use a random-walk Metropolis-Hastings step to draw θ_i given other sets of parameters (Step (9.3)). To make draws of mixture components (Step (9.4)), I use `rmultireg` and `rmixGibbs` from `bayesm` package ([49]).

ESTIMATION RESULTS

Table 10.1 and Figure 10.1 report the parameter estimates from the 2 mixture-of-normal model. There is a major group (Group 1; 84.9%) and a minor group (Group 2; 15.1%) with difference in tastes for genres, sensitivity to concert values, and the amount of prior information about the underlying concert values.

Table 10.1: Distribution of posterior means of household coefficients (θ_i): 2 normal mixture components

	Min	1Q	Median	Mean	3Q	Max
Concert feature: Time						
Weekend	-7.63	-0.17	0.31	0.27	0.77	5.22
Summer	-3.33	-0.56	-0.09	-0.07	0.41	3.88
Evening	-8.10	0.17	0.62	0.57	1.04	8.12
Concert feature: Genre						
Others	-20.65	-6.69	-5.94	-6.17	-5.25	4.28
Casual	-15.03	-2.81	-1.98	-1.78	-1.07	15.18
Chamber	-16.45	-2.65	-1.50	-1.52	-0.36	8.03
Orchestra	-9.72	-1.64	-0.70	-0.50	0.27	11.93
Family	-3.80	-2.48	-2.31	-1.11	19.20	
Jazz	-13.82	-1.42	-0.59	-0.36	0.39	13.57
Emerging professionals	-15.21	-3.96	-2.72	-2.77	-1.50	8.03
Specials	-10.36	-0.82	-0.15	-0.23	0.48	8.32
Concert feature: Price & Value						
log(Price)	-53.68	-0.03	-0.01	-0.02	-0.01	-0.00
Predicted concert value	0.006	0.661	1.441	1.398	3.025	83.72
Experience spillover						
ρ_1	-3.61	0.94	1.50	1.52	2.08	7.14
ρ_2	-6.12	-0.68	0.21	0.18	1.06	6.20
Search intensity						
ϕ_0	-9.28	-0.67	0.19	0.16	1.06	7.18
ϕ_1	-6.97	-1.53	-0.69	-0.74	0.08	4.17
ϕ_2	-6.20	-0.65	0.35	0.36	1.34	6.86
ϕ_3	-13.59	-5.60	-4.63	-4.59	-3.67	5.91
Prior knowledge						
ϕ_4	-8.78	-5.52	-4.86	-4.85	-4.18	-1.04

First, I look at the estimated prior knowledge of different groups. Figure 10.2 translates the prior knowledge parameter (ϕ_{i4}) to the fraction of the underlying concert values known to consumer i before the first visit($\phi_{ip} = \frac{exp(\phi_{i4})}{1+exp(\phi_{i4})}$; Equation 8.9). Although Group 2 is reported to have more prior knowledge about the underlying concert values with a long right
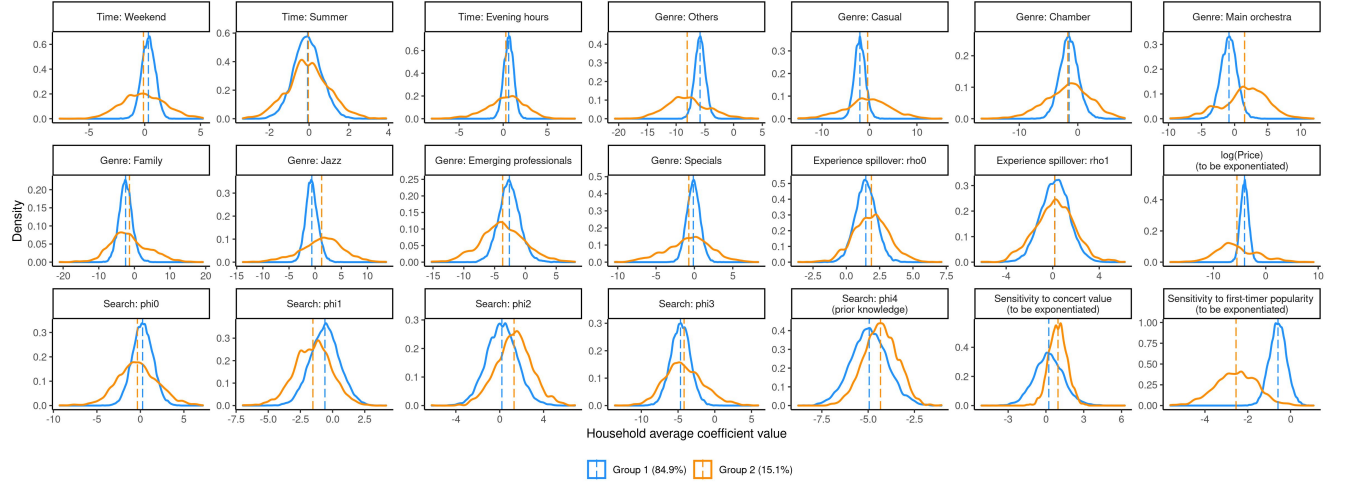
Figure 10.1: Distribution of posterior means of individual-level coefficients (θ_i): 2 normal mixture components

tail, both groups on average have little information about the overall concert values.
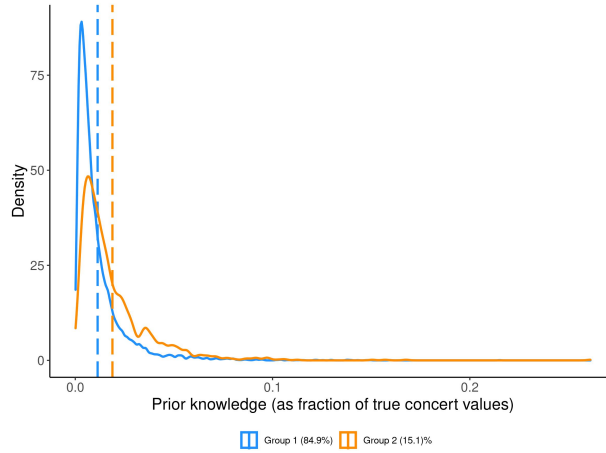
Figure 10.2: Posterior means of prior information as a fraction of true concert values: 2 normal mixture components

Figure 10.3(a) illustrates the degrees of learning spillovers at each visit. To create the empirical cdf, I take the actual observed concert choices by sample customers and their similarities to other upcoming concerts, and use them to convert the structural parameters (ρ) to the fraction of the previous concert experience that is carried over (g_{ikt} in Equation (8.4)). For example, 0.25 on the x-axis indicates that the predicted concert value of an upcoming concert is updated by $0.25 \times \Delta_{ijt}$(information discrepancy realized after visiting concert j).

60

Each line represents how learning spillovers takes place at different visits. Two patterns are noticeable. First, the average learning spillover from the initial concert value to all the upcoming concert values is 0.25. In other words, on average, if a customer comes to the symphony center for the first time with a prior belief of zero concert value, 25% of the concert value experienced at the first visit is added to the prediction of an upcoming concert value. This suggests a significant impact of a visitor's initial concert experience on her perception of all the upcoming concerts. Second, experience spillovers take place more locally as the number of visits goes up (light orange line). This implies that the same set of concerts can generate different consumer perceptions of the concert value distribution if those concerts are experienced in different orders.

(a) Learning from consumption (b) Learning from search
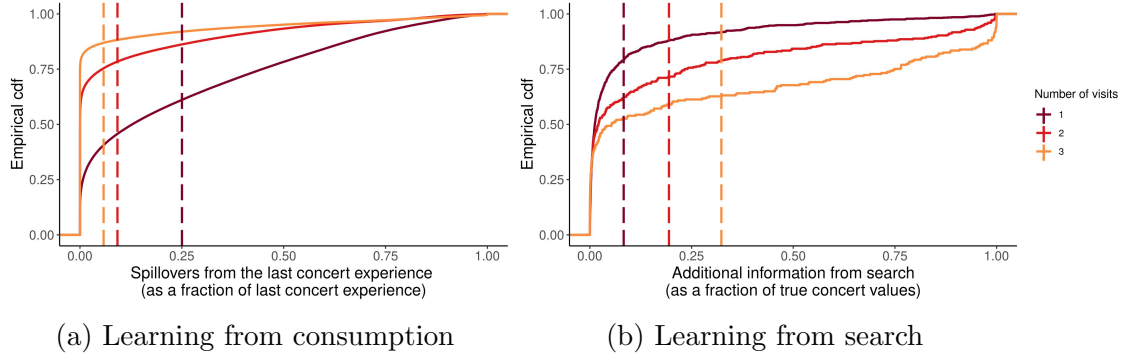
Figure 10.3: Learning spillovers from previous concert visits and additional information acquired from search at different visits

Figure 10.3(b) shows search intensity $\phi_{it} \in [0, 1]$ across visits. Here, I take the actual observed concert choices by sample customers and use them to convert the structural parameters (ϕ) to the fraction of the true convert values that is added to customer beliefs (ϕ_{it} in Equation (8.7) and (8.8)). For example, 0.25 on the x-axis indicates that the predicted concert value of an upcoming concert is updated by $0.25 \times \widehat{Q}_k^*$ ($= 0.25 \times$ the true concert value of an upcoming concert k). Each line represents how the amount of information acquired from search varies at different visits. The plot suggests that learning from search shows the opposite pattern to that of learning from previous concert experiences. First, the average search intensity is low (below 10%) after the first visit. However, it increases over visits so

61

that more than 30% of the true concert values are recovered via search ($\phi_{it} > 0.30$) after 3 visits on average.

Low prior information about the underlying concert values, high experience spillovers, and low adjustment via search after the initial visit jointly justify early customer attrition at the symphony center level. Given the lack of prior information about the concert values at the purchase stage (Figure 10.2), new consumers who randomly purchase low-value concerts in their first visits would only generalize the information (Figure 10.3(a)) without adjusting it via search (Figure 10.3(b)), which results in churn at the symphony center level.

Estimated patterns of learning underline that the order of concert visits matters in forming customers' perception about the entire concert offerings; initial concert experiences are used to update their beliefs about other concert values more globally, and the resulting beliefs are less adjusted by additional search during the early visits. This suggests that offering high-value concerts to new customers at their initial visits may have a lasting positive impact on customer retention.

Next, using the estimation results, I run counterfactual analyses to study how to effectively reverse churn induced by initial experience. I also discuss the implied trade-offs in targeted retention campaigns, price promotions, and product variety given the findings.

CHAPTER 11

COUNTERFACTUALS

Four sets of counterfactuals are run with different goals. Section 10.1 studies how to effectively reverse churn, i.e., how to re-attract consumers who have already exposed to the low-value concerts in their first visits. Section 10.2 discusses what the right target should be for the symphony center's retention efforts. Section 10.3 explores the trade-offs in price promotions when the promotions are offered on low-value concerts to clear remaining tickets, and Section 10.4 looks at the trade-offs in decreasing concert variety and its impact on ticket revenue.

11.1   Reversing churn

Here, a control group consists of consumers who experienced low-value concerts in their first visits. The objective is to re-attract these consumers, i.e., raising the average number of return visits among this specific group of consumers via marketing interventions.

Among 10000 sample customers I estimate the structural parameters for, I form a control group by selecting those customers who initially visited low-value concerts in the data. I sample those whose first concert values are lower than 10th-percentile of the concert value distribution, which results in 1509 customers. Two types of marketing interventions are applied to the group: 1) price promotion on the second visit, and 3) a hypothetical probabilistic increase in the experienced concert value at the first visit (e.g., via recommendation system or choice architecture). The second intervention, which takes place before the first visit and lowers the chance of initial exposure to low-value concerts, is presented to compare the benefit of marketing intervention before and after the visits to the low-value concerts. Each type of interventions has four levels of treatment, which is summarized in Table 11.1. Under different interventions, purchase sequence for each customer is simulated and the total number of visits is averaged across the customers over 100 weeks. 100 simulations are made
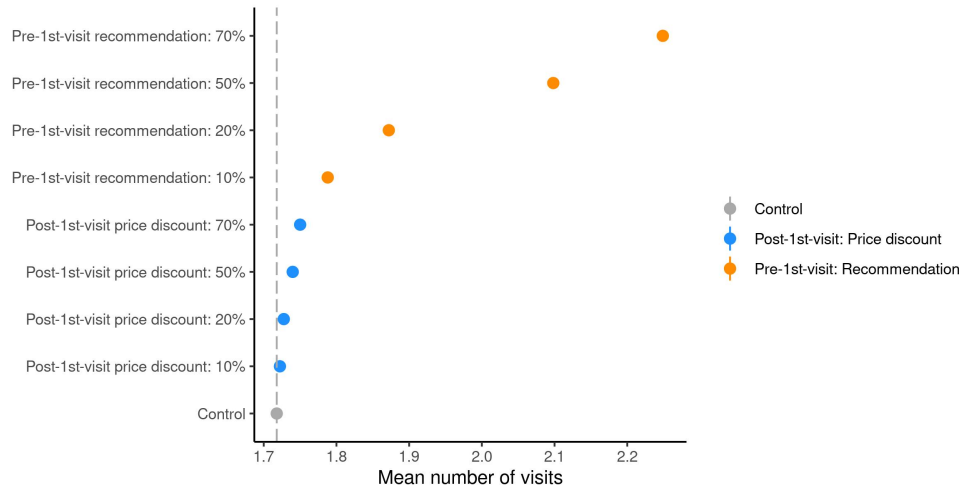
Table 11.1: Reversing churn: Counterfactual design

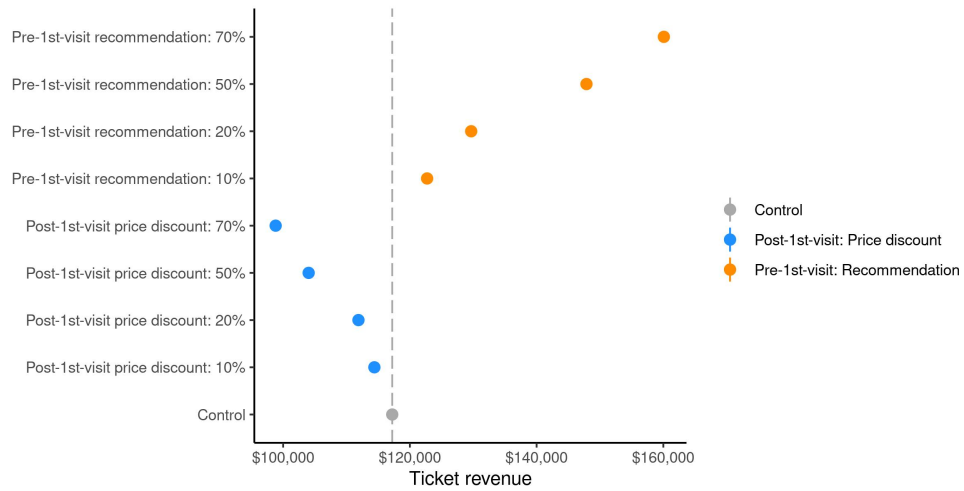| Intervention	Time of intervention	Levels	Description
Price promotion	After the first visit	10%, 20%, 50%, 70%	Offer $x\%$ of price discount on the second visit
Recommendation	Before the first visit	10%, 20%, 50%, 70%	Increase initial concert value to be the median concert value with x% chance
Control group		1509 customers who initially visited concerts with values lower than 10th percentile of the concert value distribution	

for each levels of intervention.

Figure 11.1 shows the simulation results. Post-initial-visit price discount, regardless of the size of the promotion, results in much less average return visits than the pre-initial-visit treatment ("Recommendation"). In particular, recommendation system that leads visitors to switch to a median-value concert in their first visit with 10% probability is more effective in raising the number of return visits than giving 70% price discount on the second visit is. Moreover, offering price discount even lowers the total simulated ticket revenue (Figure 11.1(b)), meaning that the loss due to price discounts is much greater than the benefit from additional return visits incurred by the discounts.

In summary, the results highlight the importance of the initial experience in customer retention given the nature of consumer learning. According to the results, the symphony center cannot re-attract consumers with price discounts in a profitable way if customers are initially exposed to low-value concert experiences. This finding underlines the importance of introductory marketing, which can be implemented via product recommendation system or well-curated choice architecture especially for prospective customers with no past experience.

Next, I explore which customers given the data on their initial concert visits should be targeted for retention efforts.

(a) Number of visits

(b) Ticket revenue

Figure 11.1: Counterfactual results: Effectiveness of different marketing interventions on reversing churn

Table 11.2: Targeted retention efforts: Counterfactual design

Intervention	Time of intervention	Levels	Description
Price promotion	After the first visit	10%, 20%, 50%, 70%	Offer x% of price discount on the second visit
Informative advertising	After the first visit	0.1, 0.2, 0.5, 0.7	Increase search intensity by x for 16 weeks after the first visit (until the second visit)
Control group		10 different groups based on the initial concert values ($< $ 10th-%, $< $ 20th-%, ..., $< $ 90th-%, $< $ 100th-% of the concert value distribution)	

11.2   Effective targeting for retention

Targeting is a common practice in churn management, but little is empirically tested on the effectiveness of such targeted retention programs (See [2] for detailed discussion). Leveraging the structural estimation results, I test which customer group, defined by initial concert values they are exposed to, gives the highest return on investment in the context of retention programs.

Two types of marketing interventions are simulated after the first visit: price promotion and informative advertising campaign. Informative advertising is translated in the model as an increase in search intensity that adjusts people's perception about the upcoming concerts; in the real world setting, it can be achieved by sending more direct mails, catalogs, or e-mail advertisements. Although it is hard to link a specific amount of increase in a structural parameter (search intensity) to the intensity of real-world marketing campaigns, this limitation does not affect the key takeaway because the purpose of this particular exercise is to see which group gives the highest return on investment rather than to compare which type of interventions (price promotions vs. advertisement) gives higher net profit. I test 4 different levels for each type of intervention, which is summarized in Table 11.2.

Figure 11.2 summarizes the results from the informative advertising condition. The figure suggests that targeting customers with 40th- to 60th-percentile concert values at their initial visits gives the highest incremental ticket revenue. The result can be interpreted as twofold. On one hand, those customers who are exposed to the concerts with high values (greater
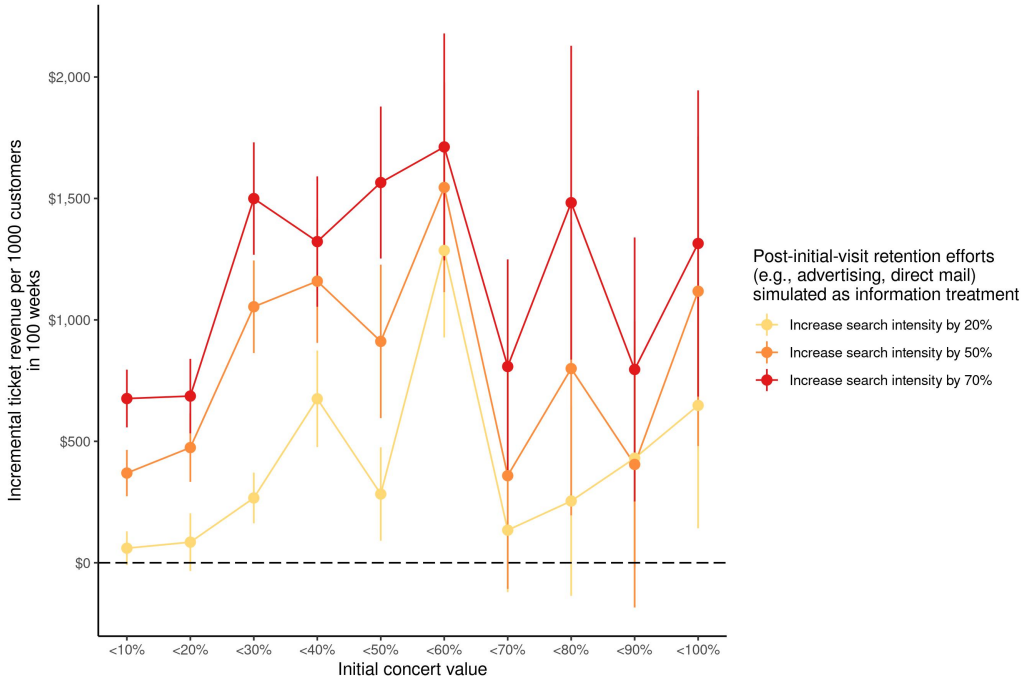
Figure 11.2: Predicted returns on informative advertising from customers with different initial concert experiences

than 60-th percentile) have enough motivation to return to the symphony center without additional marketing efforts, which makes the retention efforts less profitable. On the other hand, those customers who are exposed to the concerts with low concert values (less than 20th-percentile) would be less persuaded by advertising efforts given their strong negative beliefs carried over from the previous experience, which makes the retention efforts futile.

Figure 11.3 plots the results from the price promotion condition. Consistent with the finding in the previous counterfactual exercise (Figure 11.1(b)), price promotion results in negative profit across all conditions, suggesting that in many cases the price discounts are redeemed by customers who would have visited even without price discounts. Similar to the informative advertising case (Figure 11.2), price discounts give lower profits when they are offered to customers with initial high-value concert experiences, as those customers already have high incentives to return without price discounts. However, price promotion generates the highest profit when it is given to those customers who visited low-value concerts, which is different from the pattern found in informative advertising case.
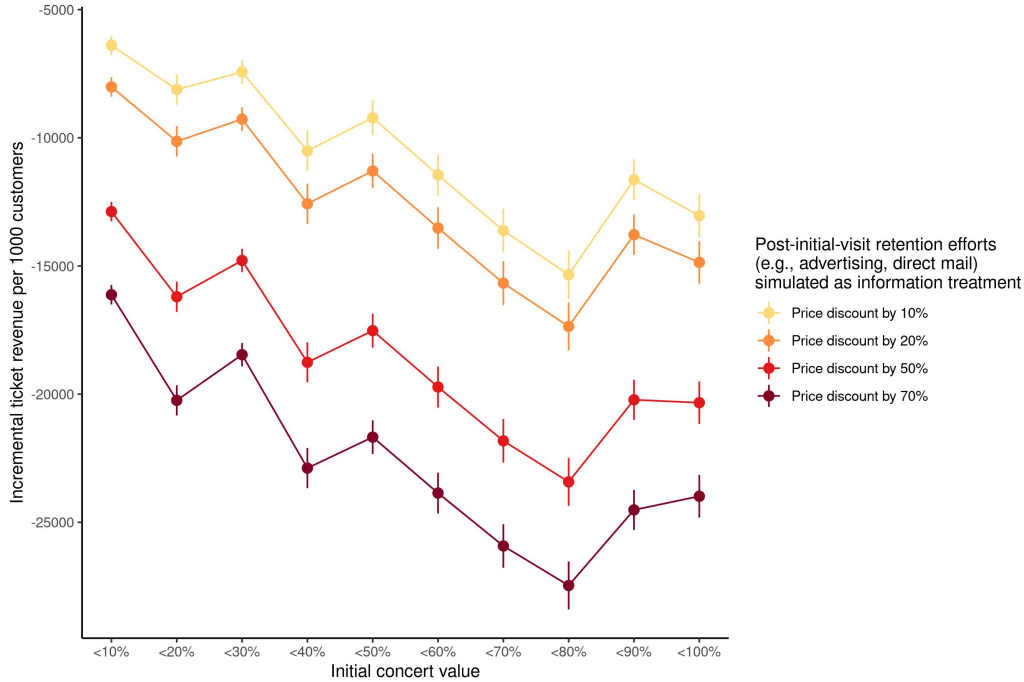
Figure 11.3: Predicted returns on price promotions from customers with different initial concert experiences

In summary, the results imply that the symphony center can increase their return on marketing efforts by selectively targeting customers based on which concerts they initially visited. Both targeted marketing efforts generate less incremental ticket revenue when they focus on customers who are already exposed to high-value experiences in their initial visits. However, comparison of the performance of advertising and price promotion highlight that the optimal target group can be different across different marketing interventions, calling for further attention from managers when designing targeted retention campaign.

Next, I explore the trade-offs in price promotions on the symphony center's ticket revenue when discounts are on low-value or high-value concerts.

11.3 Trade-offs in price promotions

Empirical findings suggest that putting price discounts on low-value items may bring unintended negative consequences on customer retention. More specifically, despite its positive

Table 11.3: Trade-offs in price promotions: Counterfactual design

Concerts on sale	Time of intervention	Levels	Description
Low-value concerts	Before the first visit (for the first visit only)	10%, 20%, 50%, 70%	Offer $x\%$ of price discount for the concerts whose values are less than 10th-%
High-value concerts	Before the first visit (for the first visit only)	10%, 20%, 50%, 70%	Offer $x\%$ of price discount for the concerts whose values are greater than 90th-%
Low-value concerts	After the first visit (for the second visit only)	10%, 20%, 50%, 70%	Offer $x\%$ of price discount for the concerts whose values are less than 10th-%
High-value concerts	After the first visit (for the second visit only)	10%, 20%, 50%, 70%	Offer $x\%$ of price discount for the concerts whose values are greater than 90th-%
Control group	2000 customers randomly selected from 10000 customers in the estimation sample		

impact on immediate sales, inventory-clearing discounts can nudge first-time customers to try low-value items and churn afterwards. In contrast, if the price promotion is put on high-value items for inexperienced customers, it may lower the immediate profit from product sales, but it may induce an increase in long-term customer retention by curating early consumption experiences to be of high-value. Other than the quality of products on sale, the time of price promotion (e.g., on the first purchase occasion, on the second purchase occasion) is also expected to play a role in customer retention, because earlier consumption experience has a larger impact on customer retention than the later experiences.

To see these potential trade-offs in price promotions, I design 2 by 2 price promotion conditions: two different sets of concerts (high-value vs. low-value) × two different times of interventions (before the first visit vs. after the first visit). Table 11.3 describes the simulated interventions.

Figure 11.4 summarizes the results. When the price promotions are offered, the average number of visits goes up across all conditions (Figure 11.4(a)). Increase in visits is more significant when the price discount is offered on high-value concerts prior to the first visit, which suggests the positive effect of nudging customers towards better initial experiences via price promotion. Despite the increase in the number of visits, the ticket revenue goes down across all conditions (Figure 11.4(b)). However, both the average number of visits

Table 11.4: Trade-offs in product variety: Counterfactual design

Intervention	Time of intervention	Levels	Description
Removing low-value concerts	Before the first visit	5%, 10%, 15%, 20%	Remove bottom xth-% concerts in the concert value distribution
Removing high-value concerts	Before the first visit	5%, 10%, 15%, 20%	Remove the top xth-% concerts in the concert value distribution
Control group		2000 customers randomly selected from 10000 customers in the estimation sample	

and the ticket revenue are higher when the price promotions are given on the high-value concerts than when they are offered on the low-value concerts (Figure 11.4(a), (c)). This again suggests that price discount on high-value concerts can create positive spillovers in subsequent revisits by guiding the initial choices of high-value concerts.

In summary, price promotions can work as an introductory marketing tool to direct first-time customers to certain items on sale, creating a lasting impact on subsequent customer-firm relationship. This implies that firms should be aware of the trade-offs between short-term and long-term profits when they organize price promotions on a specific set of items.

11.4 Trade-offs in increasing product variety

Imperfect information at the product level, combined with consumer learning at the firm level, can also create trade-offs when a firm increases its product variety. On one hand, it increases the chance of matching the tastes of broader audience, which results in higher arrival rate of customers. On the other hand, it increases the chance of customer-product mismatches by worsening the information problem at the purchase stage.

To investigate the magnitude of these trade-offs, I simulate consumer purchases for 100 weeks when some of the concert offerings are removed. Table 11.4 describes the counterfactual design.

Figure 11.5 summarizes the results. First, average number of visits drops as the number of variety goes down in most conditions because of the lower arrival rate. However, when
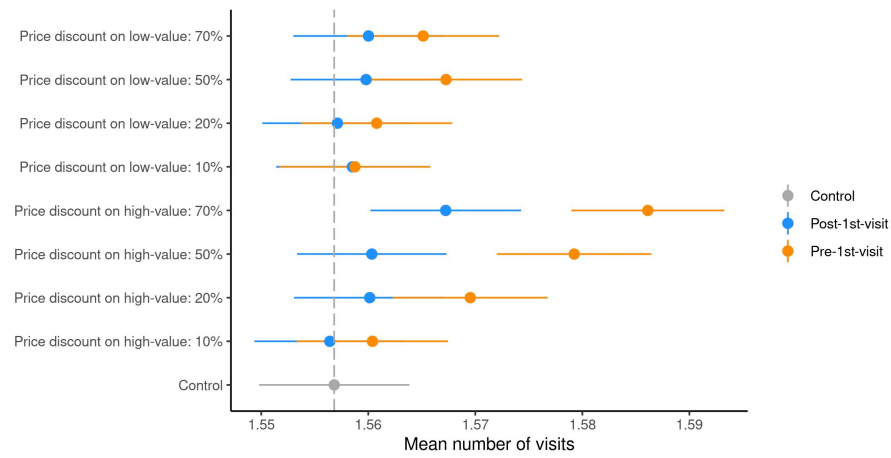
(a) Number of visits
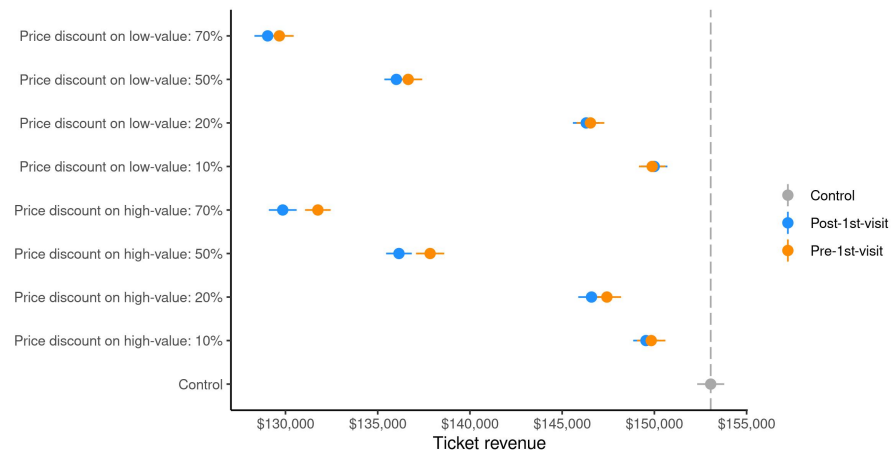
(b) Ticket revenue

(c) Ticket revenue (comparing 70% discount conditions)

Figure 11.4: Counterfactual results: Price promotions on concerts with different values

low-value concerts are removed, the mean number of visits only drops slightly or even stays statistically the same with the 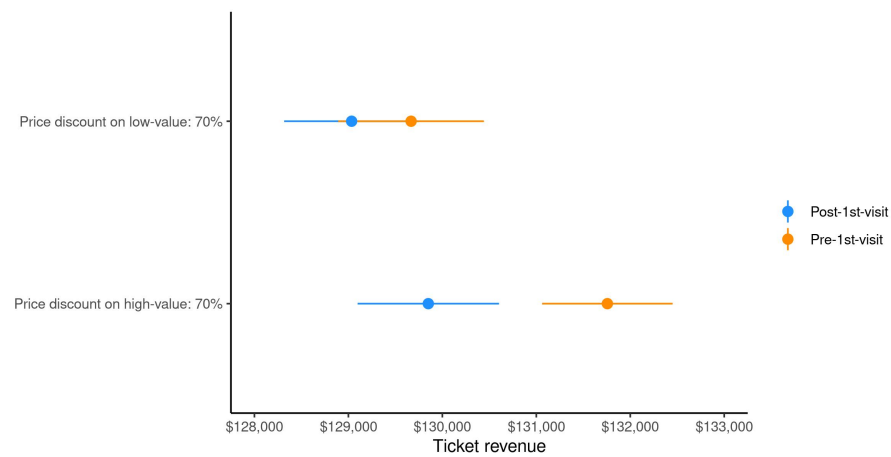number in control condition, suggesting the positive effect of reduced variety offsetting the negative effect. Second, the simulated ticket revenue is even higher when the bottom 2% concerts are dropped from the product offering. According to the result, removing the bottom 1% concerts leads to less ticket revenue than removing the bottom 2% does. This is partly because some of the bottom 1% concerts in the estimated concert values have high popularity scores among the first-time visitors; therefore, removing those concerts targeting a large group of one-time visitors may result in a more negative impact on visitor traffic. Still, since the reported ticket revenue does not take into account cost savings from staging less concerts, the net profit from concert offering reduction would be larger and positive.

Figure 11.6 shows the predicted net profit of reducing concert variety as a function of average concert production cost. To create the plot, I first compute a scale factor that matches the simulated ticket revenue in a control group to the observed market size, and multiply the simulated ticket revenues in treatment conditions by the same scale factor. Change in net profit is computed with the following equation:

$$
\begin{aligned}
\Delta\text{Net profit} &= \Delta\text{Ticket revenue} - \Delta\text{Production cost} \\
&= (\text{Ticket revenue in treatment condition} - \text{Ticket revenue in control condition}) \\
&\quad - (\#\text{Concerts in treatment condition} - \#\text{Concerts in control condition}) \\
&\qquad \cdot (\text{Av. production cost}).
\end{aligned}
\tag{11.1}
$$

To remove 1%, 2%, 3%, and 5% of the concerts in 100 weeks means to drop 4, 8, 12, and 20 concerts respectively. I calculate the change in net profit when the average production cost per concert ranges from $2,000 to $10,000. The graph indicates that removing bottom 5% concerts is predicted to create almost $120,000 of incremental profit from the entering
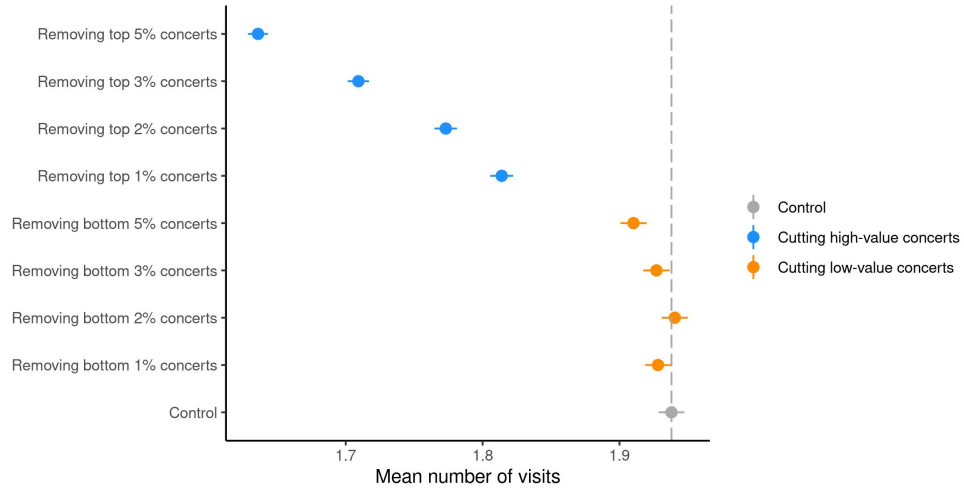
(a) Number of visits
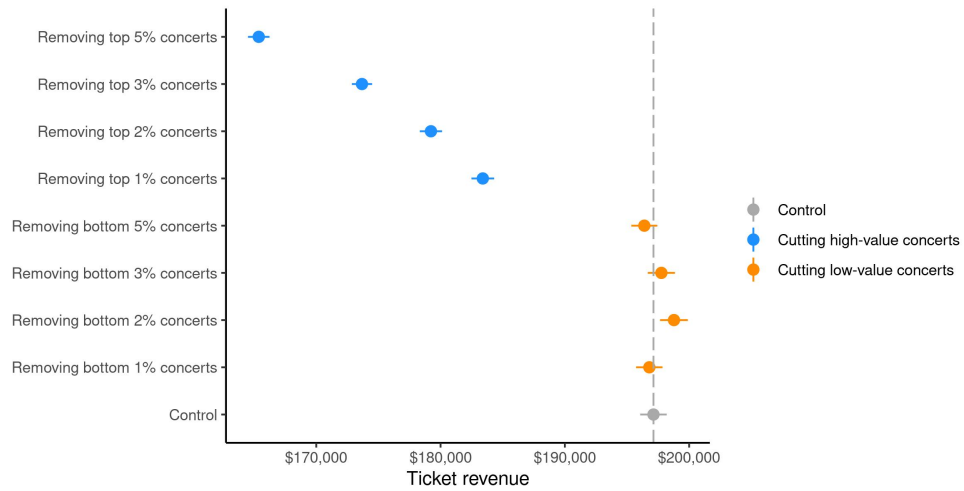
(b) Ticket revenue

Figure 11.5: Counterfactual results: Reducing concert variety before the first visit
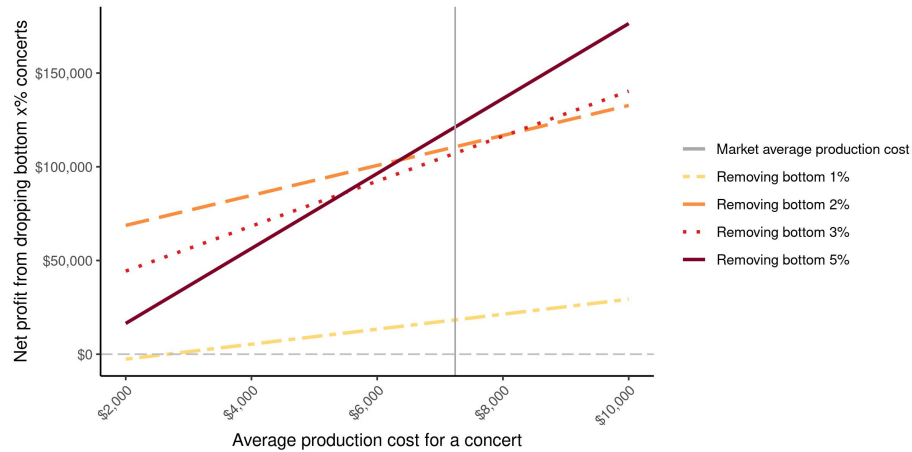
Figure 11.6: Predicted net profit from dropping low-value concerts as a function of average production cost

cohort for 100 weeks if I assume the production cost to be at the reported market average.[1] However, removing bottom 5% of the concerts will give less incremental profit than removing bottom 2% if the average production cost per concert is lower than $6,000, which highlights the tension in determining the optimal level of product variety.

---

1. From "Orchestra Facts: 2006-2014." (https://www.arts.gov/sites/default/files/Research-Art-Works-League.pdf)

CHAPTER 12

CONCLUSION

This paper shows how incomplete information at the product level can lead new customers to churn at the firm level in the context of classical music concerts. After introducing reduced-form evidence that is consistent with imperfect information, I show that two types of incomplete information jointly impact customer return decisions. First, no correlation between the arrival rates of new customers and the underlying concert values suggest that consumers are not perfectly informed about the concert values at the purchase stage. Second, a significant impact of a single concert experience on subsequent churn indicates that consumers treat a single experience to be highly representative of what the symphony center offers, suggesting the incomplete information about the range of available concert values offered by the symphony center. These two pieces of evidence explain the underlying mechanism behind why customers churn at the firm level after a single product trial; customers who buy low-value products (due to the incomplete information about the underlying product values at the purchase stage) generalize their initial experiences to all the other untried products and leave the firm (due to the incomplete information about the range of available product values offered by a given firm).

The new framework of consumer learning proposed in this paper incorporates these sources of incomplete information and rationalizes descriptive findings that are not fully justified by traditional Bayesian learning models. The framework allows consumers to have flexible patterns of learning spillovers and to have additional channel of information besides consumption. Counterfactual analyses highlight that optimal marketing strategies should explicitly consider how incomplete information affects customer retention via initial consumption experiences. More in-depth investigation on the supply-side decisions given the learning behavior may be discussed in future research. For example, understanding how to set the optimal product variety or pricing policy under a capacity constraint (e.g., only 100 seats available for each show) given the risk of early customer churn at the firm level might

be an important issue for a firm to solve.

This paper sheds light on the importance of initial experiences in customer-firm relationships, which has been documented in consumer psychology literature but not so much in empirical learning literature. While different theories and heuristics imply the potential significance of the initial consumption experience in forming subsequent actions of consumers, Bayesian learning models have not studied how the first single product experience can form customer-firm relationship onward. Moreover, no empirical studies have looked at why such a high number of churn events take place at the very early consumption stage. The gap exists partly because it is econometrically challenging to identify the causal impact of the quality of initial experience separate from an unobservable random utility component when we only observe one purchase occasion from a customer. I overcome this challenge by recovering the quality of consumption experiences (here, concert values) from experienced customers' choices and using it as data.

In summary, this paper studies why customers churn at the early consumption stage. To do so, the paper exploits rich patterns of a long individual-level panel, and proposes a new consumer learning framework to investigate optimal marketing policies for customer retention. More broadly, it opens a discussion on how consumers learn about the value of a firm based on only a few samples of experiences, although a firm these days represents a collection of diverse experiences instead of a small number of standardized products. Based on the general framework proposed in this paper, future research may extend our understanding of how imperfect information and consumer learning affects firms' optimal marketing strategies to prevent churn.

REFERENCES

[1] Eric Anderson, Song Lin, Duncan Simester, and Catherine Tucker. Harbingers of failure. *Journal of Marketing Research*, 52(5):580–592, 2015.

[2] Eva Ascarza. Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98, 2018.

[3] Eva Ascarza and Bruce G. S. Hardie. A joint model of usage and churn in contractual settings. *Marketing Science*, 32(4):570–590, 2013.

[4] Eva Ascarza, Raghuram Iyengar, and Martin Schleicher. The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, 53(1):46–60, 2016.

[5] Eva Ascarza, Scott A. Neslin, Oded Netzer, Zachery Anderson, Peter S. Fader, Sunil Gupta, Bruce G. S. Hardie, Aurélie Lemmens, Barak Libai, David Neal, Foster Provost, and Rom Schrift. In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, 5(1):65–81, Mar 2018.

[6] Eva Ascarza, Oded Netzer, and Bruce G. S. Hardie. Some customers would rather leave without saying goodbye. *Marketing Science*, 37(1):54–77, 2018.

[7] Itai Ater and Vardit Landsman. Do customers learn from experience? evidence from retail banking. *Management Science*, 59(9):2019–2035, 2013.

[8] Steven T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262, 1994.

[9] Robert C. Blattberg, Byung-Do Kim, and Scott A. Neslin. *Database Marketing*, volume 18 of *International Series in Quantitative Marketing*. Springer-Verlag New York, 1 edition, 2008.

[10] Ruth N. Bolton. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, 17(1):45–65, 1998.

[11] Ruth N. Bolton, Katherine N. Lemon, and Matthew D. Bramlett. The effect of service experiences over time on a supplier's retention of business customers. *Management Science*, 52(12):1811–1823, 2006.

[12] Sharad Borle, Peter Boatwright, Joseph B. Kadane, Joseph C. Nunes, and Shmueli Galit. The effect of product assortment changes on customer retention. *Marketing Science*, 24(4):616–622, 2005.

[13] Michael Braun and David A. Schweidel. Modeling customer lifetimes with multiple causes of churn. *Marketing Science*, 30(5):881–902, 2011.

[14] Michael Braun, David A. Schweidel, and Eli Stein. Transaction attributes and customer valuation. *Journal of Marketing Research*, 52(6):848–864, 2015.

[15] Gregorio Caetano. Neighborhood sorting and the value of public school quality. August 2016.

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

[17] Anthony J. Capraro, Susan Broniarczyk, and Rajendra K. Srivastava. Factors influencing the likelihood of customer defection: The role of consumer knowledge. *Journal of the Academy of Marketing Science*, 31(2):164–175, 2003.

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

[19] Andrew Ching, Tülin Erdem, and Michael P. Keane. Invited paper—learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013.

[20] Andrew Ching and Hyunwoo Lim. A structural model of correlated learning and late-mover advantages: The case of statins. *Management Science*, Forthcoming, 2019.

[21] Pradeep K. Chintagunta, Ronald L. Goettler, and Minki Kim. New drug diffusion when forward-looking physicians learn from patient feedback and detailing. *Journal of Marketing Research*, 49(6):807–821, 2012.

[22] Pradeep K. Chintagunta, Renna Jiang, and Ginger Z. Jin. Information, learning, and drug diffusion: The case of cox-2 inhibitors. *QME*, 7(4):399–443, Dec 2009.

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

[24] Arthur Dempster, Natalie Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, 01 1977.

[25] Jerker Denrell and James G. March. Adaptation as information restriction: The hot stove effect. *Organization Science*, 12(5):523–538, 2001.

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

[27] Larry G. Epstein. An Axiomatic Model of Non-Bayesian Updating. *The Review of Economic Studies*, 73(2):413–436, 04 2006.

[28] Tülin Erdem. An empirical analysis of umbrella branding. *Journal of Marketing Research*, 35(3):339–351, 1998.

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

[30] Peter S. Fader and Bruce G. S. Hardie. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science*, 29(1):85–93, 2010.

[31] Peter S. Fader, Bruce G. S. Hardie, and Ka Lok Lee. "counting your customers" the easy way: An alternative to the pareto/nbd model. *Marketing Science*, 24(2):275–284, 2005.

[32] Peter S. Fader, Bruce G. S. Hardie, and Jen Shang. Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29(6):1086–1108, 2010.

[33] Peter S. Fader, Bruce G.S. Hardie, and Ka Lok Lee. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4):415–430, 2005.

[34] Jr. Fryer, Roland G, Philipp Harms, and Matthew O Jackson. Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization. *Journal of the European Economic Association*, 08 2018.

[35] Kareem Haggag, Devin G Pope, Kinsey B Bryant-Lees, and Maarten W Bos. Attribution Bias in Consumer Choice. *The Review of Economic Studies*, 09 2018.

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

[37] Charles Hodgson and Gregory Lewis. You can lead a horse to water: Spatial learning and path dependence in consumer search. November 2018.

[38] Mark Israel. Services as experience goods: An empirical examination of consumer learning in automobile insurance. *American Economic Review*, 95(5):1444–1463, December 2005.

[39] Raghuram Iyengar, Asim Ansari, and Sunil Gupta. A model of consumer learning for service quality and usage. *Journal of Marketing Research*, 44(4):529–544, 2007.

[40] Sonia Jaffe, Peter Coles, Steven Levitt, and Igor Popov. Quality externalities on platforms: The case of airbnb. March 2019.

[41] Frank R. Kardes. Effects of Initial Product Judgments on Subsequent Memory-Based Judgments. *Journal of Consumer Research*, 13(1):1–11, 06 1986.

[42] Phillip Nelson. Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329, 1970.

[43] Chris Nosko and Steven Tadelis. The limits of reputation in platform markets: An empirical analysis and field experiment. February 2015.

[44] A. Yeşim Orhun, Sriram Venkataraman, and Pradeep K. Chintagunta. Impact of competition on product decisions: Movie choices of exhibitors. *Marketing Science*, 35(1):73–92, 2016.

[45] Philip M Podsakoff, Scott B MacKenzie, Jeong-Yeon Lee, and Nathan P Podsakoff. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, (5):879, 2003.

[46] Matthew Rabin and Joel L. Schrag. First Impressions Matter: A Model of Confirmatory Bias*. *The Quarterly Journal of Economics*, 114(1):37–82, 02 1999.

[47] Frederick F. Reichheld. Learning from customer defections. Harvard Business Review, March 1996.

[48] Pedro Robalo and Rei Sayag. Paying is believing: The effect of costly information on bayesian updating. *Journal of Economic Behavior & Organization*, 156:114 – 125, 2018.

[49] Peter E. Rossi, Greg M. Allenby, and Rob McCulloch. *Bayesian Statistics and Marketing*. Wiley, 1 edition, December 2005.

[50] Roland T. Rust, J. Jeffrey Inman, Jianmin Jia, and Anthony Zahorik. What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science*, 18(1):77–92, 1999.

[51] David M. Sanbonmatsu, Frank R. Kardes, David C. Houghton, Edward A. Ho, and Steven S. Posavac. Overestimating the importance of the given information in multiattribute consumer judgment. *Journal of Consumer Psychology*, 13(3):289–300, 2003.

[52] David C. Schmittlein, Donald G. Morrison, and Richard Colombo. Counting your customers: Who are they and what will they do next? *Management Science*, 33(1):1–24, 1987.

[53] David A. Schweidel, Young-Hoon Park, and Zainab Jamal. A multiactivity latent attrition model for customer base analysis. *Marketing Science*, 33(2):273–286, 2014.

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

[55] Duncan I. Simester, Catherine E. Tucker, and Clair Yang. The surprising breadth of harbingers of failure. *Journal of Marketing Research*, 56(6):1034–1049, 2019.

[56] Karthik Sridhar, Ram Bezawada, and Minakshi Trivedi. Investigating the drivers of consumer cross-category learning for new products using multiple data sets. *Marketing Science*, 31(4):668–688, 2012.

[57] S. Sriram and Pradeep K. Chintagunta. Learning models. volume 6 of *Review of Marketing Research*, pages 63–83. Emerald Group Publishing Limited, January 2010.

[58] S. Sriram, Pradeep K. Chintagunta, and Puneet Manchanda. Service quality variability and termination behavior. *Management Science*, 61(11):2739–2759, 2015.

[59] Joseph E. Stiglitz. Chapter 13 imperfect information in the product market. volume 1 of *Handbook of Industrial Organization*, pages 769 – 847. Elsevier, 1989.

[60] Maciej Szymanowski and Els Gijsbrechts. Consumption-based cross-brand learning: Are private labels really private? *Journal of Marketing Research*, 49(2):231–246, 2012.

[61] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[62] Edward Vul, Noah Goodman, Thomas L. Griffiths, and Joshua B. Tenenbaum. One and done? optimal decisions from very few samples. *Cognitive Science*, 38(4):599–637, 2014.

[63] Arthur White and Thomas Murphy. Bayeslca: An r package for bayesian latent class analysis. *Journal of Statistical Software, Articles*, 61(13):1–28, 2014.

[64] Robert S Wyer, Thomas K Srull, and Sallie Gordon. The effects of predicting a person's behavior on subsequent trait judgments. *Journal of Experimental Social Psychology*, 20(1):29 – 46, 1984.

[65] Yao Zhang, Eric T. Bradlow, and Dylan S. Small. Predicting customer value using clumpiness: From rfm to rfmc. *Marketing Science*, 34(2):195–208, 2015.

APPENDIX

A.1 Implication of the specification of $\widehat{Q}^*$: Heterogeneous preferences over genres that are known to consumers

Inclusion of genre fixed effects in Equation 5.1 allows that genre preferences are already known to consumers instead of being learned. More specifically, constructing a concert value measure as a score within a genre has two implications. First, it allows that there can be systematic differences in tastes for genres between experienced and inexperienced visitors, as the measure is net of experienced customers' baseline preferences for different genres. Second, it assumes that customers have full information about their preferences for each genre, and that there is a component to be learned for each concert within a genre. For example, all chamber music concerts can be ranked relative to one another based on their average match value to the market (which is to be learned), but the ranking is created separately for the chamber music concerts and for Jazz concerts. The choice of genre is fully driven by consumer tastes that are known to consumers from the beginning.

Here, researchers make choices on which product features are included in hidden consumption value to be learned and which are included in known preferences. In the empirical context of this paper, I categorize genre preference as known to visitors and all the other feature preferences as unknown because the first-timers' arrival rate to different concerts is uniform *within* any genre according to the data. Robustness checks with a concert value measure without genre fixed effects confirms that the major descriptive patterns stay qualitatively the same but with much more noise. Future research may investigate how to determine the structure of known and unknown preferences in a more informed way.

A.2 Validity of the estimated concert values $\widehat{Q}^*$

To check the validity of the estimated concert values, I compare the concert value estimates with the popularity of featured artists, composers, and pieces performed using the Billboard charts data. I scrape the information on top 20 classical music albums in Billboard's weekly chart during the data period and create a Billboard score for artists, composers, and musical pieces based on the their frequency of appearances on the charts. [1]. Next, for each concert I create a concert-specific Billboard index by summing the Billboard scores if any of the composers, artists, or pieces on Billboard charts are staged in a particular concert. Figure A.2.1 shows that the estimated concert values are positively correlated with such billboard scores, especially among those concerts with high estimated concert values.
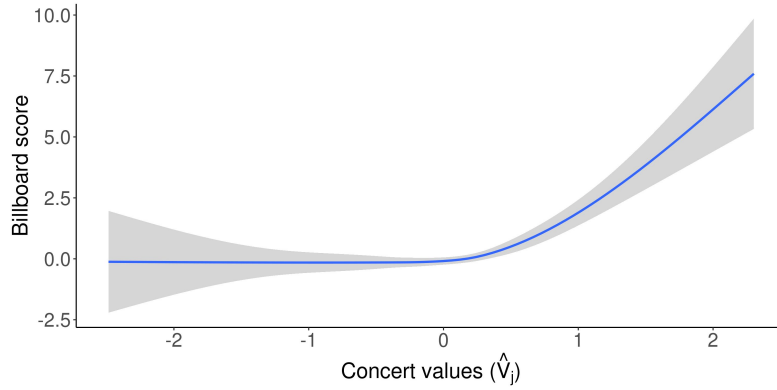
Figure A.2.1: Correlation between the estimated concert values $\widehat{Q}^*_j$ and Billboard rankings

For further validity check, I run the following regression:

$$Q^*_j \text{ chosen by consumer } i = \text{Individual FE}$$
$$+ \beta_1 \log(\#\text{Days between the performance and the ticket order date})$$
$$+ \beta_2 \log(\text{Price paid}) + \beta_3 \mathbb{1}\{\text{Purchased Seat Quantity} = 1\} + \epsilon_{ij}.$$
$$(A.2.1)$$

If the estimated concert values are valid, it should have a positive relationship with how in

---

1. https://www.billboard.com/charts/classical-albums

advance the ticket is purchased before the actual performance date (β_1). Also, it should be positively correlated with how much consumers are willing to pay for the tickets (β_2). Note that the paid price is different from the listed price by the symphony center; for example, although the listed price by the symphony center is exactly the same for two concerts, consumers still can pay different prices based on which seats they select into or whether any price discounts are offered by the venue. Therefore, the paid price represents consumers' willingness to pay for a given ticket. Finally, if the concert is of high value, there would be a set of informed customers who are willing to visit the concerts even by themselves without any company. As a result, an indicator of whether or not the quantity of tickets purchased is 1 is expected to have a positive correlation with the estimated product values. Table A.2.1 shows that all coefficients are positive as predicted. I only include non-bundle ticket purchases in this regression, but the result stays the same when I also include bundle purchases.

Table A.2.1: Correlation between the estimated concert values and other variables in the individual ticket purchase data

	Dependent variable:
	$\widehat{Q}_j^*$
log(Days of wait + 1)	0.073***
	(0.002)
log(Price paid)	0.048***
	(0.003)
$\mathbb{1}\{\text{Purchased Seat Quantity} = 1\}$	0.125***
	(0.009)
Observations	201,616
R^2	0.120
Note:	*p<0.1; **p<0.05; ***p<0.01

Finally, I check whether the estimated concert values favor specific genres or niche tastes of classical music connoisseurs. One could argue that there might be systematic difference in tastes between the experienced concert-goers and the first-time visitors. This concern is partly alleviated because the measure ranks concerts for each genre separately; if the systematic difference between the two groups of consumers arises due to the varying tastes

over different genres, the concert value used here is not affected by it. Table A.2.2 shows the list of top ranked concerts according to the concert value estimates. The list suggests that the concert value measure is neither in favor of niche tastes nor biased against the concerts targeting the mass public.

Table A.2.2: Top ranked concerts according to the estimated concert value measure

Rank	Performance name	Category name
1	Verdi	Specials
2	Oberion Trio	Chamber
3	Dudamel/Yo-Yo Ma	Specials
4	Alexandre Tharaud	Guest Piano
5	Rachmaninov 3	Main
6	Gershwin	Specials
7	Youth in Music	Emerging professionals, fusion
8	Simon Bolivar	Specials
9	2015 Festival	Emerging professionals, fusion
10	Mendelssohn Elijah	Specials
11	Kodo	Non-western
12	LOUIS	Specials
13	Lincoln Bicentennia	Specials
14	Silk Road	Non-western
15	Viva Brazil: Ma	Non-western
16	Silk Road Ensemble	Non-western
17	Movies: Williams	Movies
18	Big Green Meadow	Family
19	Brass quintet	Specials
20	P.D.Q. Bach	Specials
21	Beethoven 8 and 5	Guest Orchestra
31	Mozart Requiem	Main
32	Aladdin	Family
34	Verdi	Main
35	JaLCO	Jazz
36	Max Raabe	Specials
37	DeJohnette/Spalding	Jazz
38	Beethoven 1 and 7	Main
41	Singin' in the Rain	Movies
43	Beethoven 2 and 3	Main
44	The Firebird	Family
45	City to Country	Family
46	Rei Hotoda	Emerging professionals

A.3   Incorporating heterogeneity into the estimator of concert values

The estimator can be easily modified to explicitly allow for heterogeneity in perceived product values. One way to incorporate heterogeneity is to first cluster experienced consumers based on their purchases and to construct a separate product value measure for each customer

group. Table A.3.3 illustrates the approach.

Table A.3.3: Clustering experienced consumers based on purchase decisions

Consumer ID	Concert 1	Concert 2	...	Concert $J-1$	Concert J	Cluster
A	1	1	...	0	0	1
B	1	1	...	0	0	1
C	1	0	...	0	0	1
D	1	1	...	0	0	1
E	0	1	...	1	1	2
F	0	1	...	1	1	2
Total share	67%	83%	...	33%	33%	Average preferences of the entire market
Cluster 1 share	100%	75%	...	0	0	Preferences of Cluster 1 customers
Cluster 2 share	0	100%	...	100%	100%	Preferences of Cluster 2 customers

Here, clustering six consumers into two segments - {A, B, C, D} and {E, F} - gives two sets of within-cluster market shares that are different from the total share. Although total market shares can be used to recover average market preferences for individual concerts, within-cluster shares can offer richer information on heterogeneity in preferences over different concerts.

To recover subgroups within the experienced consumer panel, I perform latent class analysis using EM algorithm (White and Murphy 2014).

Let $Y = (Y_1, \ldots, Y_N)$ denote a binary vector of concert purchases by N experienced customers where $Y_i = (y_{i1}, \ldots, y_{iJ})$ and J is the total number of concert offerings. Each customer belongs to one of G classes and each class represents different tastes for concerts. There are two main sets of parameters: probability that an individual belongs to a group $g \in \{1, \ldots, G\}$ (denoted by π_g) and each group's purchase probability of concert j (denoted by θ_{gj}). $\pi_g \geq 0$ and $\sum_g \pi_g = 1$, and $p(y_{ij}|\theta_{gj}) = \theta_{gj}^{y_{ij}}(1 - \theta_{gj})^{1-y_{ij}}$. Purchase observations are assumed to be conditionally independent given the group membership.

The likelihood of individual i's purchase sequence Y_i can be written as

$$p(Y_i|\theta, \pi) = \sum_{g=1}^{G} \pi_g p(Y_i|\theta_g) = \sum_{g=1}^{G} \pi_g \prod_{j=1}^{J} p(y_{ij}|\theta_{gj}).$$

Let $G_i = (c_{i1}, \ldots, c_{iG})$ is a binary vector that represents i's true group membership;

$c_{ig} = 1$ if i's membership is $g \in \{1, \ldots, G\}$ and 0 otherwise. If G_i is observed with Y_i, I can write the likelihood of (Y_i, G_i) to be

$$p(Y_i, G_i | \theta, \pi) = \prod_{g=1}^{G} (\pi_g p(Y_i | \theta_g))^{c_{ig}}.$$

Since G_i is not observed, the probability for the class membership of consumer i given the observed purchase sequence is

$$p(G_i | Y_i, \theta, \pi) = \prod_{g=1}^{G} \left(\frac{\pi_g p(Y_i | \theta_g))}{\sum_{l=1}^{G} \pi_l p(Y_i | \theta_l)} \right)^{c_{ig}}.$$

I use EM algorithm to estimate θ and π. The estimation proceeds in the following step:

1. Set initial draws for θ and π and label them as $\theta^{(0)}$ and $\pi^{(0)}$. Set $k = 0$.

2. **E-step:** Update the group membership variable for each individual i $(G_i = (c_{i1}, \ldots, c_{iG}))$.

$$c_{ig}^{(k+1)} = \frac{\pi_g^{(k)} p(Y_i | \theta_g^{(k)})}{\sum_{l=1}^{G} \pi_l^{(k)} p(Y_i | \theta_l^{(k)})}$$

3. **M-step:** Update group-specific purchase probabilities and group membership probability:

$$\theta_{gj}^{(k+1)} = \frac{\sum_{i=1}^{N} y_{ij} c_{ig}^{(k+1)}}{\sum_{i=1}^{N} c_{ig}^{(k+1)}}$$

$$\pi_g^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} c_{ig}^{(k+1)}$$

4. Repeat step 2 and 3 until $\theta^{(k+1)}$ and $\pi^{(k+1)}$ converge.

$\widehat{\theta}$ and $\widehat{\pi}$ for 2 latent groups $(G = 2)$ estimated via this algorithm are presented in Figure A.3.2.
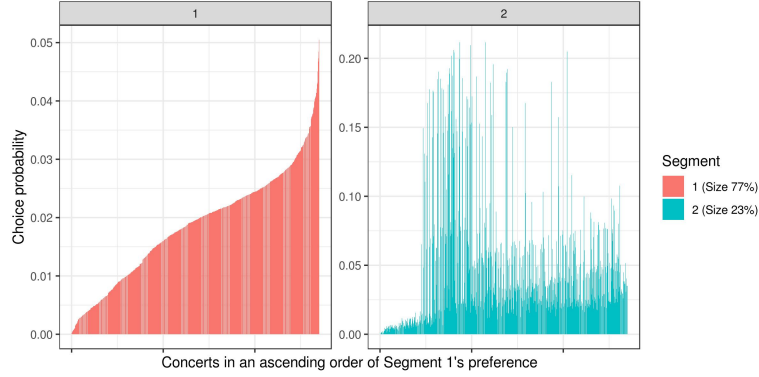
Figure A.3.2: Estimated concert choice probabilities of experienced customers by latent group

Figure A.3.2 shows the estimated choice probabilities of individual concerts of two latent customer groups. The x-axis denotes individual concert in an ascending order of Group 1's preference, and the y-axis represents the choice probability of each concert by group. Two patterns are noticeable. First, there is a big asymmetry in size between Group 1 and Group 2, which implies that the majority of consumers are clustered as the same latent class. Second, both groups agree on the concerts they do not prefer according to the low choice probabilities of the concerts located on the left side of the x-axis.

Similarly, potential systematic difference in tastes between long-tenure and short-tenure customers can be estimated separately by using different groups of sample customers. For example, in addition to a group of customers with more than 15 visits in the past, we can add another group of sample customers with 2 visits in the past and estimate the concert values using their 3rd and 4th visits only. In summary, there are many ways to extend the approach to include rich heterogeneity in preferences, which can be explored in future research.

A.4   The effect of experienced concert value on customer churn: Linear probability model with more control variables

As a robustness check, I estimate the following linear probability model:

$$Churn_{i\tau} = b_{q1}\widehat{Q}_{i\tau} + b_{q2}\widehat{Q}_{i\tau-1} b_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\Gamma + e_{i\tau} \qquad \text{(A.4.2)}$$

where $Churn_{i\tau} = 1$ if consumer i churns after τth visit and $Z_{i\tau}$ has three more sets of control variables in addition to (6.2).

$$
\begin{aligned}
Z_{i\tau} = \{ & \log(\text{Seat Quantity}_{i\tau}), \log(\text{\# Days purchased in advance}_{i\tau}), \log(\text{Price paid}_{i\tau}), \\
& \text{Concert hour}_{i\tau}, \text{Concert day of week}_{i\tau}, \text{Concert month}_{i\tau}, \text{Concert genre}_{i\tau}, \\
& \text{Popularity of the concert among the first-time customers}_{i\tau}, \text{Zip code}_i, \\
& \text{Seat Area}_{i\tau}, \text{Promotion Type}_{i\tau}, \text{Ticket Sales Channel}_{i\tau} \}. \qquad \text{(A.4.3)}
\end{aligned}
$$

The last three variables are added in the linear probability model specification. Each variable contains 59, 99, and 29 levels respectively, and controls for the effects of consumer demographics and concert experiences.

Figure A.4.3 plots the smoothed conditional mean of the residual probability of churn ($e_{i\tau}$ in Equation (A.4.2)) with respect to the initial concert values experienced. Even after controlling for a richer set of confounders, the negative correlation stays robust, which further corroborates the impact of experienced concert value on subsequent churn decision.

A.5 The effect of experienced concert value on customer lifetime value

To see if there is any additional effect of initial concert experience on customer lifetime value, I run the same regression as in Equation (6.1) using two alternative outcome variables:
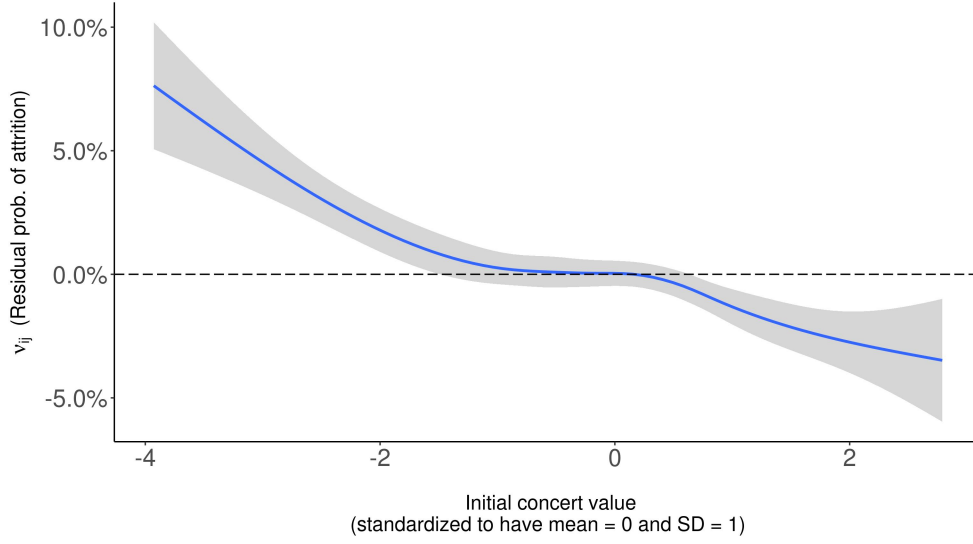
Figure A.4.3: Experienced concert value at the first visit vs. Residual probability of churn after the first visit (Linear probability model)

conversion to a regular customer (with 10+ visits) and conversion to a donor.[2] Table A.5.4 summarizes the result.

Coefficients on the experienced concert value $(\widehat{Q}_{i\tau})$ are significant and positive when the outcome variable is whether a customer stays for more than 10 visits (Column 1 to 3). As in Table 6.3, the effect of the most recent concert value persists even when the previous concert value is controlled for. Initial consumption experience also shows a significant and positive correlation with conversion to a donor (Column 4 to 6). However, the correlation becomes insignificant for visitors after three visits.

Table A.5.4 also reports the predicted changes in the outcome variables under 1-standard-deviation increase in most recently experienced concert value $(\widehat{Q}_{i\tau})$.[3] 1-standard-deviation increase in initially experienced concert value leads to $253K increase in ticket revenues over 7 years for a given consumer cohort. This estimate is much larger than the predicted increase in ticket revenue when assuming no change in churn rates beyond the second visit (Table 6.4).

---

2. I use a data set on donation activities that shares a unified customer ID system with the ticket purchase data set.

3. To monetize contribution that changes in initial concert visits can make, I use the following equation

Table A.5.4: The effect of consumption experience on customer lifetime value (Binary logit model)

	Dependent variable:					
	1 if a consumer stays for 10+ visits			1 if a consumer becomes a donor		
	(1) After 1st visit	(2) After 2nd visit	(3) After 3rd visit	(4) After 1st visit	(5) After 2nd visit	(6) After 3rd visit
$\widehat{Q}_{i\tau}$ (normalized)	0.167***	0.100***	0.114***	0.047***	0.042**	0.015
	(0.015)	(0.019)	(0.022)	(0.010)	(0.025)	(0.020)
$\widehat{Q}_{i\tau-1}$ (normalized)		0.061***	0.065***		0.007	0.030
		(0.017)	(0.021)		(0.018)	(0.025)
$\widehat{Q}_{i\tau-2}$ (normalized)			0.052***			−0.013
			(0.018)			(0.023)
log(Price paid)	0.007	-0.012	−0.051***	0.110***	0.096***	0.139***
	(0.012)	(0.016)	(0.018)	(0.013)	(0.019)	(0.023)
Observations	96,370	30,252	17,555	94,274	28,365	15,724
Log-likelihood	-18485	-11688	-9184	-23152	-10506	-6677
Controls	Y	Y	Y	Y	Y	Y
Change in outcome variable with 1 SD increase in $\widehat{Q}_{i\tau}$	5.7% → 6.7%	16.5% → 17.9%	26.9% → 28.3%	7.1% → 7.5%	13.1% → 13.6%	16.3% → 16.5%
Predicted increase in revenue/donation for one consumer cohort over 7 years with 1 SD increase in $\widehat{Q}_{i\tau}$	$253,104	$151,114	$108,108	$33,886	$16,273	$4,969

Note: *p<0.1; **p<0.05; ***p<0.01. Controls include zip code, time, and genre fixed effects. The results are robust without any control variables.

In summary, initial concert experience not just affects the probability of returning for the second visit only but also increases the probability of being a regular customer beyond the second visit. Moreover, it also has an impact on customers' subsequent donation behavior. The results implies that overall effect of initial concert experience on customer lifetime value would be large enough to make significant contribution to the symphony center's profit.

A.6   Rationalizing the descriptive findings with traditional Bayesian learning framework

Visitors' low return rate to the symphony center after a single concert visit is hard to be explained by Bayesian learning models with forward-looking behavior. In the context that this paper focuses on, a firm (the symphony center) offers a set of diverse experiences (concerts). Any rational forward-looking consumer would need at least two samples to resolve the uncertainty about the distribution of available experiences. Bayesian learning model with forward-looking behavior would not rationalize consumers not returning to the symphony center after a single visit, unless the trial cost is enormously high or the informational value of any trial becomes extremely low after the first purchase. If any of these two scenarios holds, then the motivation for modeling forward-looking behavior disappears.

Next, I discuss whether a myopic Bayesian learning model can rationalize the descriptive findings outlined in Section 5 and the potential challenges it may have.

---

to compute the predicted increase in ticket revenue and donation from 2009 consumer cohort for 7 years:

ΔTicket Revenue = Δ(Probability of being a regular consumer)

$\times$ (Average total ticket price paid by regular consumers who make first visits in 2009

- Average total ticket price paid by non-regular consumers who make first visits in 2009)

$\times$ # consumers who make first visits in 2009

ΔDonation = Δ(Probability of being a donor)

$\times$ (Average total donation made by consumers who make first visits in 2009)

$\times$ # consumers who make first visits in 2009.

Using median total ticket price or donation generates predictions that are about 60% of the reported predictions in Table A.5.4.

Suppose we use a Bayesian learning model in which consumers learn about the brand of interest (the symphony center) with normally distributed brand-level (symphony-level) signals and priors. A signal $\mu_{it}^s \sim N(\mu, \sigma_s^2)$ is drawn whenever consumer i visits a concert at time t; I assume this distribution to be common across individuals for simplicity. Consumer i has a prior on the quality of the symphony center's offerings ($N(\mu_{i0}, \sigma_0^2)$) before making any purchase. The expected utility from visiting a concert has i's expectation at time t (μ_{it}) and a random component:

$$U_{i1t} = \mu_{it} + \epsilon_{i1t}$$

$$U_{i0t} = u_{i0} + \epsilon_{i0t}$$

where ϵ_{i1t} follows i.i.d. Type 1 Extreme Value distribution and u_{i0} is normalized to be 0. if i decides to visit a concert and realizes a quality draw μ_{it}^s, then she updates her belief about the symphony center using the following equation:

$$\mu_{it+1} = \frac{\sigma_s^2}{\sigma_{it}^2 + \sigma_s^2}\mu_{it} + \frac{\sigma_{it}^2}{\sigma_{it}^2 + \sigma_s^2}\mu_{it}^s$$

$$\frac{1}{\sigma_{it+1}^2} = \frac{1}{\sigma_{it}^2} + \frac{1}{\sigma_{i0}^2}.$$

Using this simple framework, I illustrate how each data pattern is (or is hard to be) rationalized by a traditional brand-level learning model.

1) Customers do not return to the symphony center after a single visit. High churn rate after a single visit can be explained by a brand-level learning model in two different ways: 1) although prior mean was moderate, the mean of the updated belief after the first visit is low enough so that U_{i1t} never exceeds U_{i0t} (which I call 'learning' scenario), or 2) consumer prior already has low mean and variance (which I call 'preference heterogeneity' scenario). Since I rule out in Section 4 that the data pattern is purely driven by preference heterogeneity, I focus on the first case. To make the model rationalize the learning case, not

only the distribution of the signals should have high enough variance to allow for extremely negative signals, but also the weight on the signal draw should be high enough to make the updated mean belief sufficiently low after receiving a single (very negative) signal. In other words, if researchers are to explain the high churn rate with a traditional brand-level learning model, they should set both the weight on the signal draws (which is the ratio between prior and signal variance) and the absolute size of the signal variance to be high. The size of signal variance, however, has not received much attention in the literature, since many empirical works on learning normalize either signal or prior variance.

Figure A.6.4 shows an illustrative example. In each simulation, μ_0 and μ_s are both set to be 0, and the ratio of prior variance to signal variance is set to be 2:1. Seed is fixed so that the random utility component generated for each period (ϵ_{i0t} and ϵ_{i1t} for all t) are the same across different conditions. Blue solid line traces the net utility and dots indicate when purchase takes place. The four cells illustrate that it is the high signal variance case that explains churn at the symphony level after a few visits.

Although capable of replicating the high churn rate with certain parameter values, the brand-level learning model still cannot allow consumers to endogenously select the next brand-level signal, as the model assumes that a signal is randomly drawn at each purchase. This is restrictive in the empirical context of the paper in which customers select the symphony-level signal by making concert choices.

2) Customers become better at choosing high-value concerts even among concerts with new features that they never experienced before. Any learning model that explains this pattern should assume correlated learning across different products or attributes. In addition to the intractability of this framework given the large number of available concerts and concert features, it is challenging for the existing correlated model to explain that customers choose high-value concerts even if those concerts do not necessarily share the features with previously visited concerts.
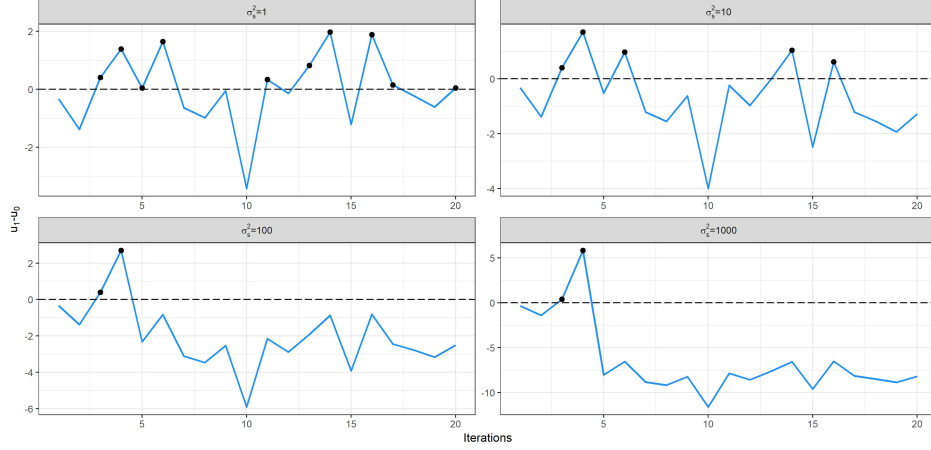
94

Figure A.6.4: Rationalizing the data pattern using traditional Bayesian learning frmaework: Simulation results ($U_{i1t} - U_{i0t}$)

A.7   Computing concert similarities

Two different approaches are used to construct concert similarities given a large set of binary concert features:

- *Logistic Principal Component Analysis (PCA)* (Langradf and Lee 2015): I first run Logistic Principal Component Analysis on 615 binary concert features representing 1350 concerts, setting the number of principal components to be 100 (whose result explains 98.1% of the concert feature variance). Using the resulting 100 components that are continuous, I calculate the euclidean distance between any given pair of concerts.

- *Gower's metric* (Gower 1971): Gower's metric is defined as $G_{ij} = \sum_k d_{ijk}$ where d_{ijk} is the difference between concert i and j in feature k. In the weighted Gower's metric which assigns different weights to binary features, I define the weight attached to each concert feature based on the total number of appearances of the specific feature across different concerts. For example, the genre feature "Chamber" appears much more frequently than the composer feature "Schubert" because there are more concerts that belong to Chamber series than the concerts staging Schubert pieces. As genre captures more fundamental differences in concerts, "Chamber" is assigned with greater weight
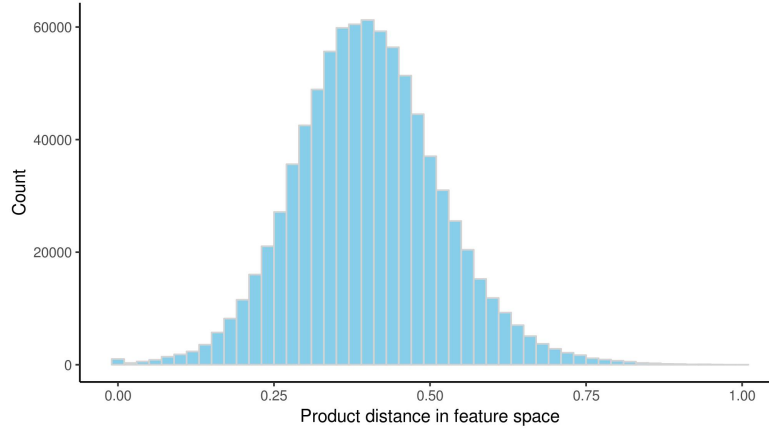
Figure A.7.5: Distribution of concert similarities created with Gower's metric (normalized to be between 0 and 1)

than "Schubert" based on the frequency of feature appearances.

Figure A.7.5 shows the distribution of concert similarities created with Gower's metric. Concert similarities are normalized such that they lie between 0 and 1.

A.8 Correlation between concert prices and concert values $(\widehat{Q}^*)$

Figure A.8.6 shows the correlation between the listed prices and the estimated concert values within genre. As the plot shows, the correlation between concert values and the list prices is not strong. Some genres show no correlation (e.g., Emerging professionals) because every concert in this specific category is offered at the same price.
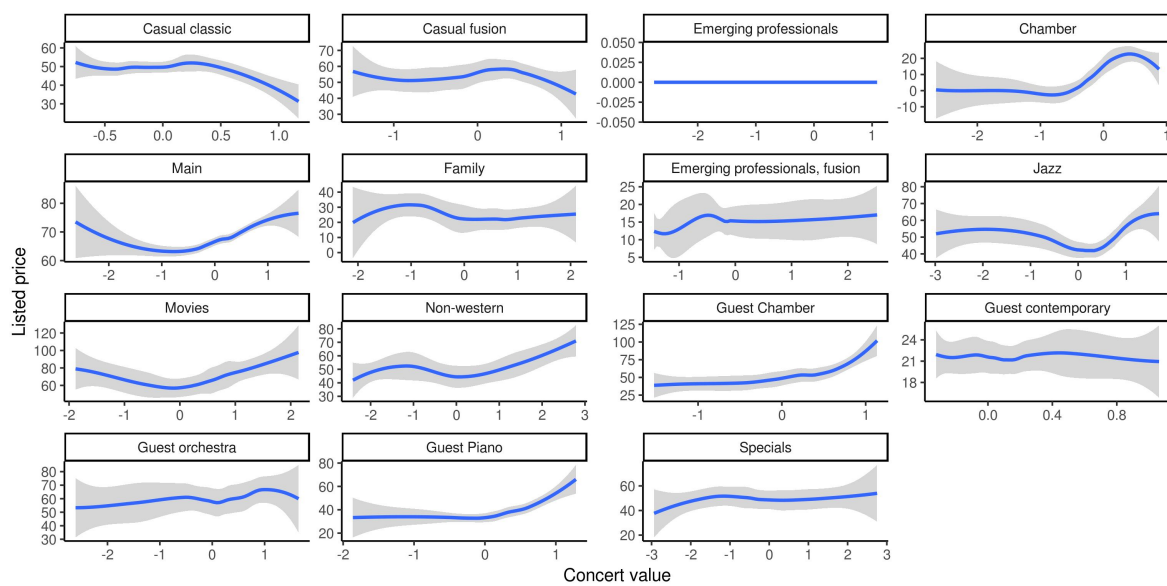
Figure A.8.6: Correlation between concert values and concert prices