Customer retention under imperfect information

Yewon Kim*

August 2019
Preliminary and in progress

Abstract

I study why many firms face low retention rates among *new* customers. In particular, I study whether firm abandonment after a single product experience is solely driven by heterogeneous preferences or is affected by information frictions about the products. I use a novel, long panel of individual-level ticket purchases from a major U.S. symphony center for which 60% of first-time customers do not return. The data exhibit patterns consistent with learning and information frictions in the market. Descriptive analyses document information frictions and learning spillover that jointly cause customer attrition. First, many customers attend concerts with a low match value due to their incomplete information about concert qualities. Second, a low match value at the initial visit creates a strong adverse learning spillover effect by reducing 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 information frictions. Through counterfactual analyses, I analyze both a policy that offers high-quality concerts to first-time customers and a policy that offers targeted pricing to second-time customers after their initial visit to low-quality concerts. The results emphasize the importance of introductory marketing to new customers.

keywords: Customer retention, Brand, Information, Churn, Consumer learning, Experience goods, Product Variety.

*Ph.D. student in Marketing, University of Chicago Booth School of Business, yewonkim@chicagobooth.edu. I am grateful to my committee, Sanjog Misra, Bradley Shapiro, Jean-Pierre Dubé, and Sarah Moshary, for their unlimited support. I thank Pradeep Chintagunta, Anita Rao, Günter Hitsch, Øystein Daljord, and the participants at the Chicago Booth Marketing student seminars for their helpful feedback. I also thank the data provider for all the resources. I am responsible for all remaining errors.

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 including for meal kits, hotels, and food delivery service.[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 consumers' true preferences who have full information about the firm's 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 study whether customer churn at the firm-level is solely driven by heterogeneous preferences or is also affected by imperfect information about available products in the firm. I use a novel, long panel of individual-level ticket purchases from a major U.S. symphony center for which 60% of first-time consumers do not return at least for 4 years.[2] The data contain 13-year-long individual-level 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 full information about the underlying product values.[3] The test I develop takes the learning models that have been widely used in the marketing literature and extracts a testable prediction from those models, which therefore documents *reduced-form* evidence of consumer learning. Under a standard learning model in which consumer belief about an unknown quality converges to its true value, experienced consumers' choices should reflect unobserved 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, *within* customers, new customers' product choices converge towards the experienced customers' choices over time. The pattern of within-customer choice convergence is not explained under a model that assumes fully informed, fixed heterogeneous preferences. This test does not require a specific functional form for the learning rule, such as Bayesian updating rule, and therefore can be used to test more general patterns of consumer learning.

[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.

[3] Product values can contain vertical quality, horizontal match value, or both.

Next, based on the finding that learning exists, I derive and estimate concert values (i.e., average preference perceived by the market) as a function of experienced consumer choices. The evidence of consumer learning suggests that experienced consumers are better informed about the underlying product values, and therefore their choices contain information about true product values. The estimated measure of product values allows me to document consumers' lack of information about the product values at the initial purchase stage. In addition, it allows me to trace how consumers learn over time without assuming a specific learning rue by observing how consumers make choices in terms of product values over time.

Descriptive analyses show how imperfect information causes customer attrition in two stages. First, many new customers attend low-average-match-value concerts due to their incomplete information about the concert values at the initial purchase stage. The data show that the arrival rate among the new consumers is *random* with respect to the underlying concert values revealed by experienced consumers' choices. Second, a low-match-value experience at the initial visit has a significant impact on the probability of subsequent churn, even after controlling for potential heterogeneous preferences among the first-time customers, which indicates a strong spillover effect of a single concert experience to consumers' perception of all future concerts. In other words, they treat a single experience as representative of what the symphony center offers, although the product they consumed lies in the low tail of match value distribution. The pattern of strong spillover from the *initial* product experience 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 a firm 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 other data patterns that are not fully captured by traditional Bayesian 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 entire symphony center's offering 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 even when they face new concert features they have not tried before. To reflect these patterns, I also allow consumers to acquire information through an additional channel besides consumption, which I call information-seeking behavior. In this model, consumer decision to engage in information-seeking is determined by how pleasant the previous product experiences were, which creates an incremental impact of prior product experiences on subsequent purchase behavior. The model suggests that a pleasant prior experience increases not only the likelihood of returning to the firm but also the likelihood of purchasing high-value products in the next period due to the information-seeking triggered by the positive experience in the previous period.

The key parameters of the estimated structural model reflect 1) how far consumers generalize the sample information from a single concert experience to other concerts untried, and 2) how consumers obtain additional information on concert values via information-seeking based on the quality of their previous experiences. The estimated parameters give numerical interpretations on the lasting impact of few consumption experiences on consumer tenure under imperfect information.

Counterfactual analyses explore the importance of introductory marketing to new customers to steer them towards better first-time experience. The results show that even 50% discount for the second visit is not sufficient to match the effect of high-quality initial experience on the number of return visits. The results also emphasize the potential trade-offs a brand faces when it sets optimal marketing strategies given the specific learning behavior. For example, in the case of price promotions, putting low-match-value products on sales to clear inventories may have a negative effect on customer acquisition because inexperienced consumers who are nudged to buy products on discounts may lower their expectation about the brand based on the specific low-match-value product they purchased on sale.

This paper complements the existing literature on the drivers of customer retention by applying consumer learning theories. Consumer churn as a topic has been extensively discussed in marketing literature (Schmittlein et al. (1987), Fader et al. (2005a), Fader et al. (2005b), Zhang et al. (2015), Ascarza & Hardie (2013), Ascarza, Netzer, & Hardie (2018), Capraro et al. (2003), Iyengar et al. (2007), Sriram et al. (2015)). However, a large amount of effort has been put to predict when customers churn, and there have been surprisingly few studies that explain *why* consumers do not return (Ascarza (2018)). Moreover, to my knoweldge, there is no empirical study that looks at why such a high number of churns take place at the early stage, although most statistical models that predict churns fully take into account this pattern 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 churns based on customer needs and learning pattern (Iyengar et al. (2007), Sriram et al. (2015), Nosko & Tadelis (2015), Ascarza et al. (2016), Ascarza, Neslin, et al. (2018)). 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 various disciplines implies the potential welfare loss caused by information frictions (Nelson (1970), Stiglitz (1989), Denrell & March (2001), Israel (2005)).

The paper complements the literature on consumer learning. It shows the 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 spillover as well as incorporates an additional endogenous information channel besides consumption. It also illustrates how the initial consumption experience - which is assumed to be "debiased" with more purchases in Bayesian learning literature - can be deterministic of the future consumption and other information acquisition path, as documented in consumer psychology literature (Tversky & Kahneman (1974), Kardes (1986)) and few empirical studies (Ater & Landsman (2013), Haggag et al. (2018)). More broadly, the paper sheds light on 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 sampling one of them. This implies that the signals consumers receive and the construct they learn about no longer align perfectly, and that experience spillover via correlated learning may take place very broadly as indicated in previous studies on correlated learning (Erdem (1998), Coscelli & Shum (2004), Sridhar et al. (2012), Szymanowski & Gijsbrechts (2012), Che et al. (2015), Ching & Lim (2019)).

The findings underline the potential trade-offs a firm would face when making marketing decisions given the specific learning behavior. For example, incomplete information at the purchase stage, combined with strong spillover from the initial experience, suggests that enriching product variety can have two opposing effects on firm's profit: 1) it can increase the profit by satisfying the tastes of broader audience, and 2) it can decrease the profit by raising the probability of any new consumer choosing a low-match-value product and leaving the brand afterwards. Similarly, price promotions put on low-quality products can result in two diverging impacts: 1) it can increase the profit by clearing inventories, and 2) it can decrease the profit by nudging new consumers to first try low-value products and to update their perception about all the other untried products accordingly. These trade-offs suggest that optimal marketing strategies should explicitly consider that customers may have information frictions when making product choices as well as when updating their perception subsequently. The findings also imply that a firm should pay more attention to informing its new customers instead of focusing only on its loyal customers (Rust et al. (1999)).

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 (Orhun et al. (2016), Caetano (2016)), it is new to my knoweldge to recover product qualities using the choices of a consumer subgroup based on consumer learning theory. 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 (Podsakoff et al. (2003)), and industries also question the validity of the survey results based on their low predictive power (Reichheld (1996)). The proposed methodology would allow researchers to control for unobservable features without exploiting any other data sets.

The rest of the paper proceeds as follows. Section 2 discusses the related literature. Section 3 introduces the general setting of interest and summarizes the data. Section 4 presents a test for information frictions and consumer learning. Section 5 discusses the identification of product qualities from purchase data only. Section 6 shows the descriptive evidence that information frictions increase customer attrition, and Section 7 proposes a framework of consumer learning that justifies the findings. Section 8 and 9 discuss how the model is specified and estimated, and Section 10 reports the results. Section 11 discusses counterfactual analyses, and Section 12 concludes.

2 Literature Review

In this section, I explain how this paper is related to the literatures 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 (Blattberg et al. (2008)). Research on customer attrition has studied two different settings: a contractual setting in which consumer departure is observable by the 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 the aggregate pattern of consumer churns using various variables ranging from consumer demographics to consumption history (Schmittlein et al. (1987), Fader et al. (2005a), Fader et al. (2005b), Zhang et al. (2015), Ascarza & Hardie (2013), Ascarza, Netzer, & Hardie (2018)). Much less discussion has been made on understanding the causal drivers of customer churn (Bolton (1998), Borle et al. (2005), Braun & Schweidel (2011), Sriram et al. (2015), Ascarza (2018)). Moreover, to my knowledge, although many models that predict churns fully take into account the commonly observed pattern of high churn rates at the early stage of consumption when fitting the data, there is no empirical study that focuses on why the churn rate is higher among first-time customers and whether there can be any improvement in retention at the early 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 model with rich consumer decision theories. Using a micro-founded consumer utility model instead of a statistical 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 customer needs and learning pattern (Iyengar et al. (2007)). Also, the results suggest that the high attrition rate among new customers should not be assumed to be a market feature that is uncontrollable but be viewed as a measure to be improved.

2.2 Bayesian learning models

Researchers have developed different models to reflect how consumers learn about the underying qualities of experience goods via purchases (Nelson (1970)). 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 [Chintagunta et al. (2009), Chintagunta et al. (2012), See Sriram & Chintagunta (2010) for review] vs. forward-looking [Erdem & Keane (1996), See Ching et al. (2013) for review]). Models also vary in the scope of information spillover. Information from one purchase occasion can have spillover effects across different categories, brands, or product attributes (Erdem (1998), Coscelli & Shum (2004), Sridhar et al. (2012), Szymanowski & Gijsbrechts (2012), Che et al. (2015), Ching & Lim (2019)).

Acknowledging the limitations of strong assumptions imposed by Bayesian learning, 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 (Epstein (2006), Rabin & Schrag (1999)), allows consumers to revise signals or prior beliefs in a subjective manner. (Fryer et al. (2018), Camacho et al. (2011))

Despite the richness of modeling approaches, empirical literature on consumer learning faces several challenges. First, most papers do not test whether or not learning actually drives consumer purchases before proposing rich learning models that support the observed data pattern. Identification of most learning models is based on functional form assumptions (e.g., which prior information consumers have, how consumers learn and update their expectation) (Ching et al. (2013)), and only few papers document the evidence of (no) learning in a more model-free (or less model-heavy) way (Chintagunta et al. (2012), Dubé et al. (2010)). Modeling consumer learning based on the assumption that learning occurs may make the results less interpretable since it is hard to disentangle learning from preference heterogeneity in many empirical settings (Shin et al. (2012)). Second, most papers do not consider that brand-level learning is based on a product-level signal, and that the product-level signal a consumer receives is an outcome of her purchase decisions. Prior research on learning assumes that consumers receive a random signal about a brand from any product purchase, not allowing the possibility that each (product-level) signal consumers get might not be fully random depending on which product they choose. Third, most papers do not model how product experience from previous consumption changes consumers' subsequent information acquisition process (Heilman et al. (2000)). Although consumers engage in both search and consumption to obtain different types of information about products or categories, there has been no empirical paper to my knowledge that explicitly models how these two information acquisition activities interact with each other and jointly affect consumer purchase decisions.

This paper extends the existing literature in several ways. First, I show the 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 information frictions 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 the signal from one product experience to affect the perception of all the other untried products in a tractable way. Third, I jointly model consumers' purchase and search decisions. In my model, consumers can get information from search as well as from consumption, and the decision to search is endogenously determined by the information acquired from previous consumption.

2.3 Consumer judgment given limited information

Consumer psychology literature documents that consumers' initial judgment given limited information may not only be more extreme (Sanbonmatsu et al. (2003)) but also be hard to be adjusted by subsequent stimuli (Tversky & Kahneman (1974), Kardes (1986)). The paper by Sanbonmatsu et al. (2003) 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 brand's product with limited information about the brand, their judgment about the brand 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 and put more certainty on it than experienced consumer would.

The lasting impact of the initial signals can also be explained by anchoring and adjustment heuristics (Tversky & Kahneman (1974)) or the initial judgment effect (Kardes (1986)). Anchoring followed by insufficient adjustments suggests that different initial signals can result in different estimates about the same value, which are biased toward the starting points (Tversky & Kahneman (1974)). The initial judgment effect, first documented in the context of social judgments (Wyer et al. (1984)), indicates that initial judgments of a target affect subsequent memory-based judgments of the same target (Kardes (1986)). Memory-based judgments based on the initial judgments are also correlated with people's interest in seeking more information about the target product (Kardes (1986)). This implies that the initial negative signal can block further information channel, which makes the impact of the negative signal on subsequent purchase behavior even stronger.

3 Setting

Section 3 introduces the general setting of interest and the specific empirical setting the paper focuses on. In particular, it discusses how imperfect information can be a driver of customer attrition in a market with multiple types of information frictions.

3.1 General setting - A market with imperfect information

I focus on the market with two sources of information frictions. The first source is that consumers do not have full information about the underlying product values (qualities or match values or both) until they consume. Examples include any markets with experience goods (Nelson (1970)). The second source is that consumers do not have full information on available products or available product qualities a brand offers. Examples include markets for clothing items, furniture, stationary products, and food items (e.g,. cereals), in all of which each brand provides more than a handful of choice alternatives. Consumers engage in different information acquisition activities to reduce each type of frictions. In response to incomplete information on the underlying value of a given product, consumers make purchases to learn about the value a firm's product. Facing a large set of available items from a given firm, consumers engage in additional information-seeking activity prior to any purchase to learn about the values of the firm's other items they have not consumed before. I call this information channel "search."[4]

Each of the two information acquisition activities - search and purchase - results in a signal that consumers use to update their beliefs about the products of interest. In particular, beliefs carried over from previous consumption experience determine the amount of subsequent search activities as well as subsequent purchase decisions. For example, if the past product experience was positive, then consumers would search more products of the same firm in the next shopping period, would find a high match-value product with higher probability, and would have another pleasant experience from the brand. In contrast, if the previous purchase resulted in low satisfaction, consumers may turn off their incentives to search the firm's other untried items, which decreases the probability of finding the high value products from the firm even if they exist.

Figure 1 illustrates how prior product experiences shape the following 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 the items available at the moment(J_t). Her decision to search is based on the trade-off between the expected gain and cost from search. Expected search benefit depends on the perceived mean and variance of the untried product qualities, both of which are learned from the past purchases.

Let B_t denote the belief about product values at time t that has been obtained from the past product experiences. Let S_t denote the information on the 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

[4]Here, search does not mean any specific means of online or offline search but refers to any information acquisition behavior, ranging from paying attention to the content of the brochure to reading online reviews.

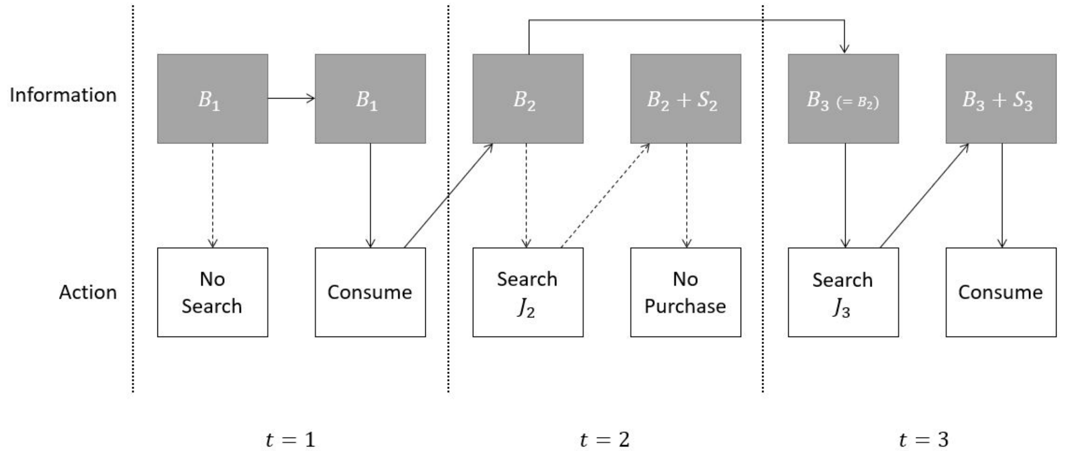


Figure 1: Consumer information acquisition process

information set is $B_t + S_t$ if she has engaged in search, and B_t if she has not.

In this framework, prior product experience affects the next period's consumer information in two ways. First, it directly enters the information set based on which the purchase decision is made. Second, it affects the amount of additional information obtained through search (S_t), which also enters the information set. These direct and indirect effects of the experienced product values can generate a snowball effect on customer retention. On one hand, positive consumption signal 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. On the other hand, negative consumption signal 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 with imperfect information on product values and product availability, a single product experience can exert a lasting influence on the following customer-firm relationship 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 symphony center in the U.S. 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 contains consumer-level observations on 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.

Classical music concert consumption is a representative example of the general setting described above. 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, consumers necessarily rely on a small number of prior consumption experiences to decide their subsequent search, purchase, and return decisions. A long listing of available concerts with various programs requires consumers to use heuristics to extrapolate the information they already have when predicting the quality or match value of each concert. Third, 60% of the first-time visitors do not return to the symphony center for at least 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 the concerts in the upcoming few months 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 1 reports descriptive statistics of the ticket purchase data. The full data set contains purchases from FY2005 to FY2018, but for the analysis I use purchases from FY2009 to FY2015 (i.e,. have both burn-in and burn-out periods) to alleviate potential biases from left truncation. 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 orders made by each consumer over seven years is 1 (Table 1(b)). Higher mean and maximum of the total orders indicate that the distribution of total orders 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(≈ 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 the left truncation bias when labeling consumer 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 1(b)). These statistics are based on the prices chosen to be paid by consumers and not on the prices offered by the symphony center. A major portion of variations in price *offered* is within-concert, i.e., it comes from different seat locations within a concert (See the last column of Table 2).

Consumer demographics are 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.

Table 1: Descriptive statistics

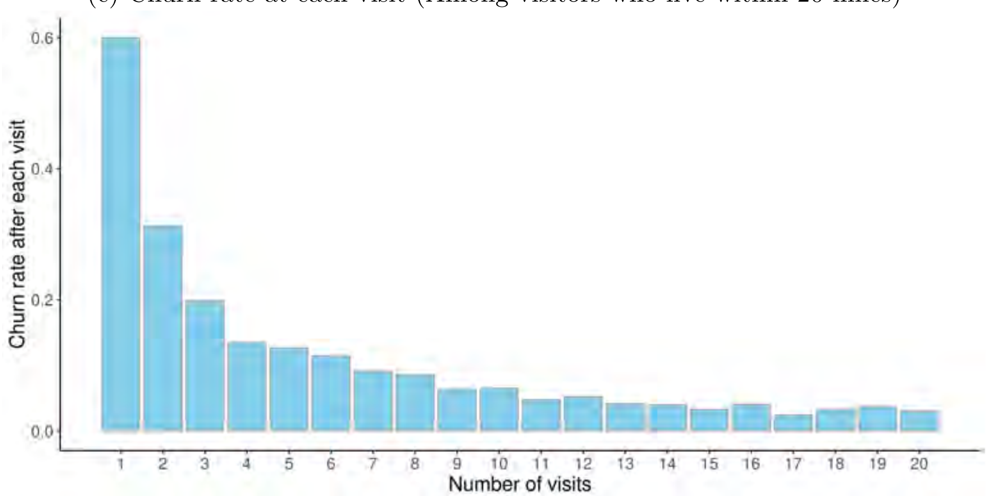
(a) By fiscal year

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	9	4	582
Days between purchases (Among those with total orders >1)	1	21	44	109	118	2449
Price paid per ticket	0	29	47	55	77	750
Within-individual SD of the price paid	0	6	13	17	24	433
Per-capita income (zip-code level)	16	28,824	38,466	44,059	55,034	200,208
Distance (zip-code level)	0	6	17	145	39	4,310

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



3.2.2 Concert features from program catalogs

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 (contemporary or classical), the types of pieces performed (Jazz, movie sound tracks, or classical pieces), musical composition (chamber, orchestra, or solo), ambiance (more casual and experimental or traditional and classic), whether the guest performers are invited, whether student artists perform, and the target audience (family-friendly or not). Each category has different baseline price, which is one of the two major sources of variations in price paid other than seat locations (Table 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.

Table 2: Category-level price (in $)

		All		Within-concert
Category	Unique concerts	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 variation comes from the difference in baseline price across categories or genres, there is still within-genre across-concert variation in price. To capture price differences across concerts, I compute the average price of the seats in Main Floor for each concert.[5] Table 4 reports the summary statistics.

[5] I use only the observations that entail full price.

Table 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

Table 4: Summary statistics: Price (in $)

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

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

4 A test for imperfect information and documentation of consumer learning

To study whether imperfect information causes customer attrition, I first construct a *test* that informs whether there is imperfect information at the purchase stage in the context of classical music concerts. This step of investigating whether imperfect information is observed in the data of interest is missing in many empirical studies that build rich learning models based on the assumption of imperfect information followed by consumer learning. The test I develop 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. If consumer beliefs are fixed over time under the model of fully informed preferences, then there should be no time-varying correlation between the choices by inexperienced consumers and those by a separate group of experienced consumers. If the correlation between the two sets of choices varies systematically over the inexperienced consumers' purchase occasions, then it implies that the beliefs change as more consumption experiences are accumulated. The pattern of changing beliefs about product values rejects the model of fully informed fixed preferences and supports the existence of information frictions at the initial purchase stage.

Given the test result that is consistent with the prediction under standard learning models, I show that the true product values 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 product values based on the invertibility condition (Berry (1994)), I document the *path* of consumer learning by tracing the estimated product values chosen by a fixed group of consumers over their purchase occasions.

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 v-th visit (purchase occasion) is

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

and her choice probability is

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

$\mathscr{S}(\alpha_i, Q_{ijv})$ is 1) everywhere differentiable w.r.t. Q_{ijv}, and 2) $\frac{\partial \mathscr{S}_j}{\partial Q_{ijv}} > 0$ & $\frac{\partial \mathscr{S}_j}{\partial Q_{ikv}} < 0 \ \forall k \neq j$ (Berry (1994)).

Let $\mathcal{E} = \{i | v \geq \overline{v}\}$ denote a sample group of experienced consumers whose number of past visits (v) is greater than or equal to $\overline{v}$. Similarly, $\mathcal{I} = \{i | v < \overline{v}\}$ 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: $Q_{ijv} = Q_{ij} \quad \forall v$), 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. $v_{\mathcal{I}}$ represents the number of visits $\mathcal{I}$ have made.

If consumer beliefs about product values are stable, neither $\boldsymbol{Q}_{\mathcal{I}}$ nor $\boldsymbol{Q}_{\mathcal{E}}$ change over purchase occasions $(\nu_{\mathcal{I}}, v_{\mathcal{I}})$, which implies the following:

$$\begin{gathered}
\boldsymbol{s}_{\mathcal{I}} = \mathscr{S}(\alpha_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{I}}) \\
\boldsymbol{s}_{\mathcal{E}} = \mathscr{S}(\alpha_{\mathcal{E}}, \boldsymbol{Q}_{\mathcal{E}}) \\
\frac{\partial \; corr(\boldsymbol{Q}_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{E}})}{\partial v_{\mathcal{I}}} = \frac{\partial \; corr(\alpha_{\mathcal{I}}, \alpha_{\mathcal{E}})}{\partial v_{\mathcal{I}}} = 0 \\
\Rightarrow \frac{\partial \; corr(\boldsymbol{s}_{\mathcal{I}}, \boldsymbol{s}_{\mathcal{E}})}{\partial v_{\mathcal{I}}} = 0
\end{gathered}$$

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,

$$Q_{ijv} \to Q_j^* \text{ as } v \to \infty.$$

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 $v < \underline{v}$.

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

Under the model of learning that satisfies A1-A3, $\boldsymbol{Q}_{\mathcal{I}}$ converges towards a vector of true product values $\boldsymbol{Q^*}$ as $\mathcal{I}$ accumulates more experiences (v).

$$\begin{gathered}
\boldsymbol{Q}_{\mathcal{I}v} \to \boldsymbol{Q^*} \approx \boldsymbol{Q_{\mathcal{E}}} \quad \text{as} \quad \nu \to \infty \quad (\because A3) \\
\boldsymbol{s}_{\mathcal{I}} = \mathscr{S}(\alpha_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{I}v}) \\
\boldsymbol{s}_{\mathcal{E}} = \mathscr{S}(\alpha_{\mathcal{E}}, \boldsymbol{Q}_{\mathcal{E}}) \approx \mathscr{S}(\alpha_{\mathcal{E}}, \boldsymbol{Q^*}) \\
\frac{\partial \; corr(\alpha_{\mathcal{I}}, \alpha_{\mathcal{E}})}{\partial v_{\mathcal{I}}} = 0, \quad \frac{\partial \; corr(\boldsymbol{Q}_{\mathcal{I}}, \boldsymbol{Q}_{\mathcal{E}})}{\partial v_{\mathcal{I}}} > 0 \\
\Rightarrow \frac{\partial \; corr(\boldsymbol{s}_{\mathcal{I}}, \boldsymbol{s}_{\mathcal{E}})}{\partial v_{\mathcal{I}}} > 0
\end{gathered}$$

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 experienced consumers' choices stays the same across the purchase occasions. i.e.,

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

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

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

To test the hypotheses, I use the following step:

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

   I create $\mathcal{E} = \{i|v \geq 15\} = \{$a sample of 5000 experienced customers who have made at least 15 past visits$\}$.[6] $s_{\mathcal{E}j}$ for concert j is calculated as

   $$s_{\mathcal{E}j} = \frac{\sum_{i\in\mathcal{E}} y_{ij}}{\sum_{i\in\mathcal{E}} \sum_{k\in\{k|Year_k = Year_j\}} y_{ik}}$$

   where $y_{ij} = 1$ if consumer i purchases tickets for concert j. The denominator is the total number of sample experienced consumers who visit the symphony center in a given year. Any transactions by consumers who belong to $\mathcal{E}$ are taken out from the data for the analysis.

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

   $$(s_{\mathcal{E}j} \text{ of concert } j \text{ chosen by consumer } i \notin \mathcal{E} \text{ at visit } v) = \alpha_i + \beta_v v + \eta_{ijt} \quad \forall i \in \mathcal{I}. \tag{1}$$

3. Check whether β_v is statistically different from 0.

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

Test results Figure 2 reports mean $s_{\mathcal{E}j}$ chosen by consumers across visits. Each line denotes the average $s_{\mathcal{E}j}$ chosen by a fixed group of consumers with a given observed length of tenure. The

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

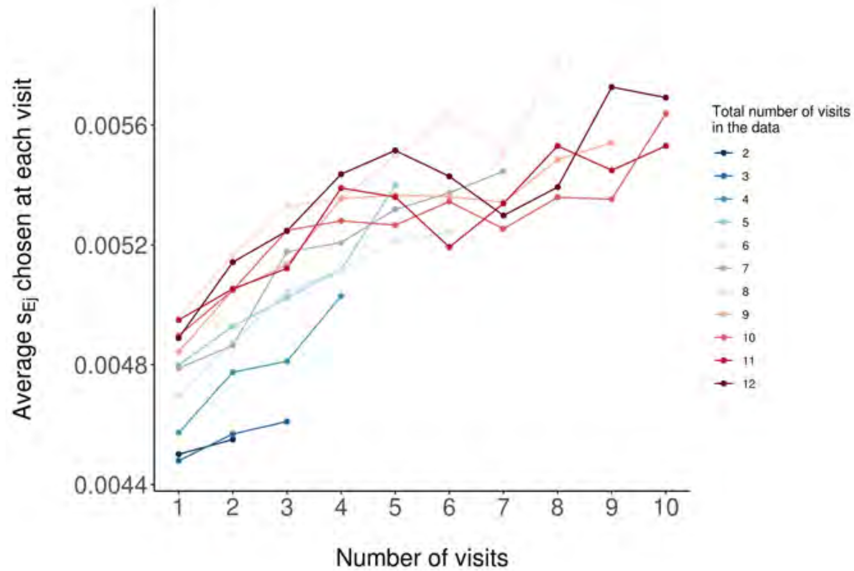


Figure 2: Correlation between the experienced and inexperienced consumers' choices over time

plots are generated using balanced panels, which means that for each line the customer base stays the same across the number of visits (x-axis). Upward-sloping patterns in both plots imply that the choices of inexperienced consumers become more concentrated around the concerts with high experienced consumers' shares ($s_{\mathcal{E}j}$) as the number of visits by inexperienced consumers goes up.

Table 5 reports the test results. The results with statistically significant β_v's reject the null hypothesis that the correlation between the experienced and inexperienced consumers' choices stays constant across inexperienced consumers' purchase occasions.

Table 5: 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 $\geq$ 10 visits
β_v (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

The test result provides the evidence that consumer beliefs about product values are *not* stable over consumption experiences. There can be various factors that drive changes in consumer beliefs over experiences. Consumer learning about product values through purchases rationalizes the pattern of changing beliefs. Targeted advertisement that promote different products to different groups of customers based on the length of consumption history can also justify the same pattern. In this specific context, the advertisement effect is ruled out given the content of the marketing materials provided by the symphony center.

Although the test does not *accept* consumer learning, it *rejects* the model of fixed preferences under perfect information (Remark 1). Also, the pattern of choice convergence towards what is more favored by experienced consumers over purchase occasions 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, as any consumer learning framework assumes imperfect information in the market.

Based on the implication from the test, I show that the underlying product values can be identified from experienced consumers' choices under the standard model of consumer learning.

4.2 Identification of product values

Given the test results that reject the model of perfect information with fixed preferences, I identify product values and document consumer learning patterns under the assumption of standard learning models (A3).

Using *A1-A3*, I show that, under the assumed model of learning, the true product values can be inverted from the observed market shares 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 choice shares of consumers with sufficiently high number of past consumption experiences ($\boldsymbol{s}_{\mathcal{E}}$).

Sketch of proof: Let $\boldsymbol{Q}_{iv}$ denote a vector of consumer i's beliefs about product values after v visits (purchase occasions). A3 implies that there is $\overline{v}_\delta$ s.t. $|\boldsymbol{Q}_{i\overline{v}_\delta} - \boldsymbol{Q}^*| < \delta \quad \forall \delta > 0$. For each $\delta > 0$, let $\mathcal{E}_\delta$ denote a set of experienced consumers $\{i|v_i \geq \overline{v}_\delta\} = \{i||\boldsymbol{Q}_i - \boldsymbol{Q}^*| < \delta\}$. 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 Berry (1994).

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

An estimator of the underlying product 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,[7] 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):

$$\begin{aligned} \widehat{Q}_j^* &= \ln(s_{\mathcal{E}j}) - \ln(s_{\mathcal{E}j0}) - \text{(Hour, Day, Month, Genre Fixed Effects)} \\ &\text{where} \quad s_{\mathcal{E}j} = \frac{\sum_{i \in \mathcal{E}} y_{ij}}{\sum_{i \in \mathcal{E}} \sum_{k \in \{k | Year_k = Year_j\}} y_{ik}} \quad \text{and} \quad s_{\mathcal{E}j0} = 1 - s_{\mathcal{E}j} \\ &\text{and} \quad \mathcal{E} = \{i | v \geq \overline{v}\}. \end{aligned} \tag{2}$$

The distribution of estimated product qualities using 5000 experienced consumers with $\overline{v} = 15$ is summarized in Figure 3.[8] These estimates, computed with the assumption of no heterogeneity, identify the *average* preferences perceived by the market. Section A.3 in the Appendix further discusses the validity of the estimated measure, in which I compare the estimate product value measures with billboard rankings of featured artists and composers and the average number of days the tickets are purchased in advance for each concert.

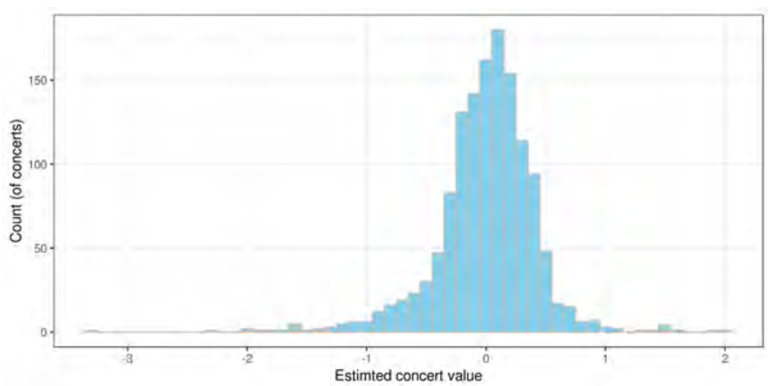


Figure 3: Histogram of quality measures across concerts

The estimator can easily be modified to explicitly account for heterogeneity in perceived product value. One approach that can be implemented - clustering experienced consumers using their purchase sequences and constructing a separate measure within each cluster - is introduced in Section A.4. For subsequent analyses, I use $\widehat{Q}_j^*$ without consumer heterogeneity (following Equation (2)) as a proxy for market average perceived quality.

To my knowledge, it is new to the field to identifying unobservable product qualities by applying the prediction of the standard learning models to the choices of a consumer subgroup. However, the idea of using choices by individuals outside the group of interest has been applied in prior research

[7]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.

[8]As in Section 4, I use different values of $\overline{v}$ from 15 to 30 for robustness check, and different cutoff values give highly correlated estimates (correlation greater than 0.94).

to control for unobservables. For instance, Orhun et al. (2016) 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, Caetano (2016) controls for neighborhood unobservables by using the residential choices of people without school-age children.

Constructing a proxy for product values using revealed preferences of informed consumers 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 (Podsakoff et al. (2003)), and industries also question the validity of survey results based on their low predictive power (Reichheld (1996)).

Using the estimated concert values, I document consumer learning patterns without assuming a specific updating rule by tracing how consumers choose concert values across purchase occasions.

4.3 Documentation of learning

Once having the product values estimated from the experienced consumers' choices, researchers can leverage them to study consumer learning patterns without assuming any specific learning rule. Investigating how fast or slowly consumers' product value choices reach the steady state can inform researchers how fast consumers learn about the underlying product values.

Figure 4 shows the learning pattern for a group of consumers who make more than 18 purchases during the data period. The x-axis denotes the number of visits, and the y-axis denotes the average value of the concert chosen at each visit. First, the upward sloping pattern suggests that consumers' choices become more concentrated around high-value concerts as they make purchases. Second, the concavity of the curve suggests that the rate of learning or updating of consumer beliefs about concert values slows down and consumers reach the stationary state for concert choices.

Figure 15 in the Appendix shows the average concert values chosen at each visit by consumer groups with different length of observed tenure. Most groups, regardless of how many total they have made during the data period, exhibit the upward sloping pattern between the number of visits and the average concert value purchased, which supports the model of consumer learning about the underlying concert values.

This approach of documenting consumer learning deviates from how most empirical learning literature shows learning. Instead of estimating consumer preferences and learning parameters based on specific functional form assumptions (e.g., Bayesian updating rule), this approach first estimates consumers' true preferences using experienced consumers' data based on a more flexible assumption and traces how consumers learn over time by observing which product values consumers choose over time. The assumption that this approach requires is that consumers' beliefs about

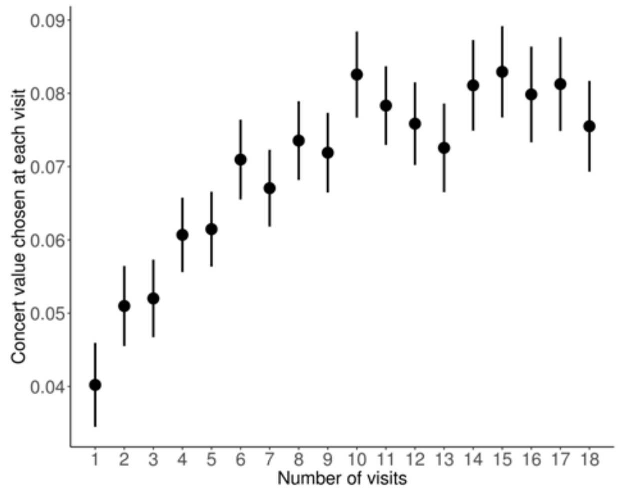


Figure 4: Choices of concert values at each visit (among consumers who make ≥ 18 visits)

product values converge to the true product values as they get more experienced, which holds for most standard learning models. As discussed in Section 4.2 and Section A.4, heterogeneity in perceived product values can be easily incorporated when estimating the product values. This new perspective complements the literature by introducing a way to relax certain assumptions about the path of learning.

Next, I run descriptive analyses using the estimated quality measure and show how imperfect information causes customer attrition.

5 Descriptive Analysis

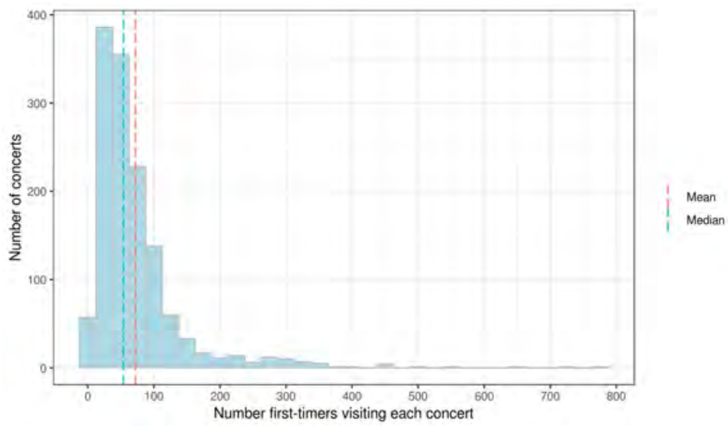
In this section, I show through descriptive evidence that initial concert visits chosen under imperfect information significantly affect customer attrition at the symphony center level. First, I document the high mean and variance of the first-time customers' churn rates after visiting different initial concerts, which indicates potentially large weight consumers put on a single experience when making return decisions. Second, I show that the arrival rate to different concerts is almost random among the first time visitors despite the variance in concert values, which suggests information frictions about the underlying concert values at the purchase stage. Finally, I show that the concert values realized at the first few visits shape consumer churn decisions.

For the descriptive analysis, I include only those households within 30 miles from the symphony center based on the zip code information.

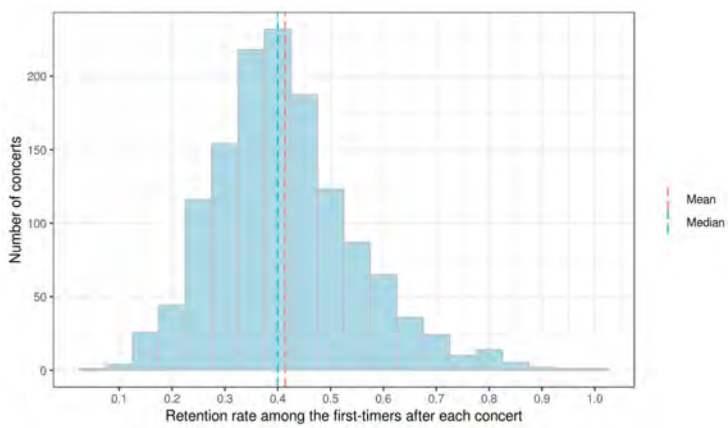
5.1 Varying retention rates across the concerts

Figure 5 shows that there is enough variation in both the number of the first-time customers who arrive at each concert (Figure 5(a)) and the rate of churn after each concert (Figure 5(b)). As

Figure 5(b) shows, the average retention rate across concerts is about 40%, implying that more than half of the visitors at any given concert do not come back to the symphony center after their first visit at least for four years. 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 timers 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 5: Distribution of the first-timers' arrivals and retention rates

5.2 Imperfect information about concert values at the purchase stage

If the variation in retention rates described in Figure 5 is purely due to consumers' fully informed preferences, any marketing strategy that inform customers about the available concerts and available concert values may not effectively raise retention. However, if at least part of the churn events are attributed to the mismatches between concerts and consumers due to the imperfect information at the purchase stage, correcting these mismatches might favor both consumers and the symphony center.

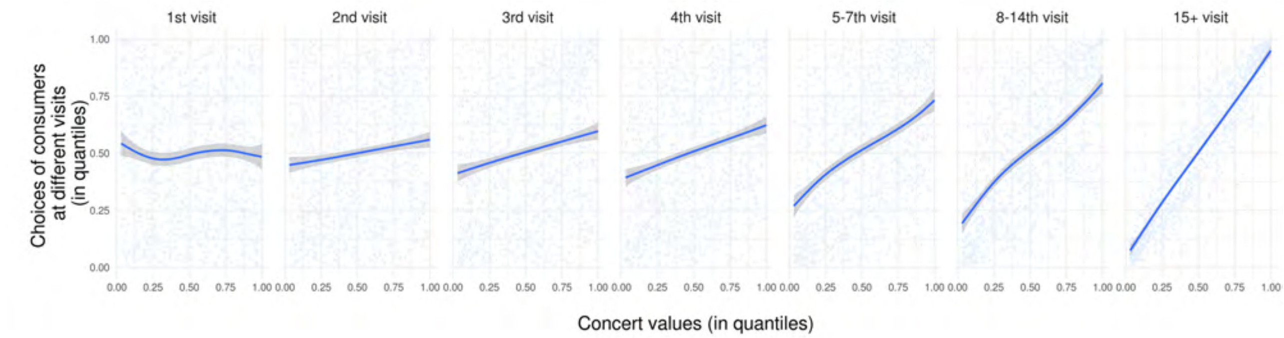


Figure 6: Concert values vs. Concert choices by consumers with different numbers of past visits: Quantile-Quantile scatterplots and smoothed conditional means

To identify if there is imperfect information at the purchase stage about concert values, I first revisit Figure 4 in Section 4.3 and Figure 15 in Section A.2. Both figures implies that the average concert value chosen in the initial visit is lower than the concert value chosen in the later visits for any consumer group, which suggests that there exist information frictions among new consumers at the purchase stage.

Table 6: 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 shows how consumers with different number of past visits choose concerts compared to the experienced consumers. The x-axis denotes the quantile of the estimated concert values $\widehat{Q}_j^*$, which is created using Equation 2 with a holdout sample of 5000 consumers with more than 15 visits. The y-axis denotes the quantile of concert choices (in log odds, net of time and genre fixed effects) for consumer group with different levels of experiences; it is constructed with concert choices made by consumers after each visit using Equation 2.

Two patterns are noticeable. First, the initial choices by the first-time customers are random with respect to the estimated concert values (the first cell of Figure 6).[9] Second, consumer choices become more concentrated around the high-value concerts as the level of experience goes up. Table 6 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 consumers get more experienced.

One can argue that the pattern of choice convergence towards high-value concerts in Figure 6 or Table 6 is solely due to the systematic change in customer base due to preference heterogeneity; consumers with low customer lifetime value might have preference for specific types of concerts that regular consumers do not seek for, which may create a similar choice resemblance pattern. However, I show in Section 4 that the data rejects the model of fully informed preferences, which undermines the plausibility that the pattern is created solely due to this alternative explanation.

In summary, there exist information frictions at the purchase stage about the true concert values among inexperienced consumers, which lead inexperienced consumers to make random choices of concerts with respect to the concert values.

5.3 The effect of experienced concert value on customer return decision

Given that consumers' first concert choices are random with respect to the concert values, I look at whether the experienced concert values affect first-time customers' returns to the symphony center.

First, to analyze the effect of the experienced concert value on subsequent churn decision, I run a regression with individual-level churn decision as an outcome variable and the recent concert value consumed as the key explanatory variable:[10]

$$Pr(Churn_{i\tau} = 1|\tau\text{-th visit}) = \frac{\exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)}{1 + \exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)} \tag{3}$$

where $\widehat{Q}_{i\tau}$ is the estimated value of the concert that i visited at her τ-th visit. Similarly, $\widehat{Q}_{i\tau-1}$ and $\widehat{Q}_{i\tau-2}$ are the lag concert value experienced at $\tau-1$-th and $\tau-2$-th visit. To ease the interpretation, for the descriptive analysis I normalize $\widehat{Q}_{i\tau}$ to have mean 0 and standard deviation of 1. I include 8 variables as control variables ($Z_{i\tau}$, Equation 4) Table 7 describes each variable.

[9] Result remains the same when I use raw scales instead of quantiles.

[10] 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 7: 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 (2) 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)

$$
\begin{aligned}
Z_{i\tau} = \{ &\log(\text{Seat Quantity}_{i\tau}), \log(\#\text{ Days purchased in advance}_{i\tau}), \log(\text{Price paid}_{i\tau}),\\
&\text{Concert hour}_{i\tau}, \text{Concert day of week}_{i\tau}, \text{Concert month}_{i\tau}, \text{Concert genre}_{i\tau},\\
&\text{Popularity of the concert among the first-time customers}_{i\tau},\\
&\text{Zip code}_i \} \qquad (4)
\end{aligned}
$$

Popularity of the concert among the first-time customers are created using Equation (2) but using the choices of the first-time customers instead of the choices of sample experienced customers. One might expect that there are certain concerts that are popular among one-time customers which are positively correlated with subsequent churns and are negatively correlated with experienced consumers tastes. If a set of such concerts exist, it will result in a negative correlation between the estimated concert values and subsequent return rates. To control for this potential spurious correlation, I include the measure of popularity among the first-time customers to absorb the systematic variations in target customers across different concerts (i.e., controlling for those concerts that are only tailored for casual visitors). Table 8 summarizes the results.

Significant and negative coefficients of concert values imply that there is a systematic correlation 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 subsequently. 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 is correlated with the reduction of the churn rate from 61.3% to 59.5%. The effect of the concert value experienced most recently does not vanish even when the lag concert value is included in the regression. If we assume that the first concert value chosen by any consumer informs her latent type (e.g., long tenure vs. short tenure), controlling for the concert value chosen at the first visit should be controlling for consumers' latent type. 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 the experienced concert value on consumer decisions to churn. 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 due to spurious correlation driven by changes in the composition of consumer demographics.

Figure 7 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 recently experienced concert value ($\widehat{Q}_{i\tau}$).[11] 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.

[11] I include upto 3 lag concert value variables in this specification.

Table 8: 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
Change in churn rate with 1 SD increase in $\widehat{Q}_{i\tau}$ (using av. churn rate)	61.3% → 59.5%	34.1% → 32.6%	22.9% → 22.2%	61.3% → 58.5%	34.1% → 31.0%	22.9% → 21.0%

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

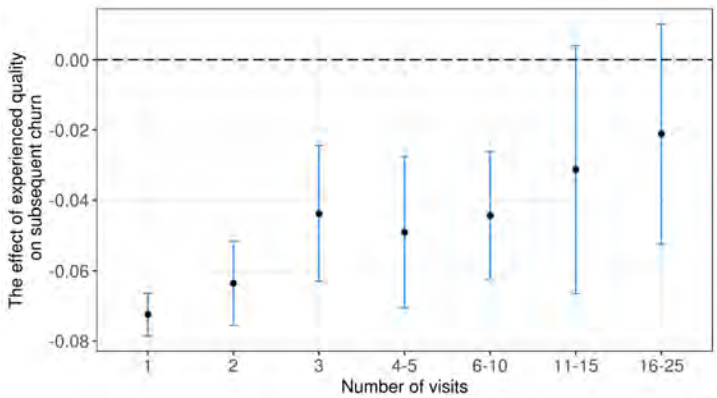


Figure 7: The effect of recent quality consumed on subsequent churn over visits

Table 9: Predicted increase in ticket revenue from 1.8% decrease in churn after the first visit (in $)

	2009	2010	2011	2012	2013	2014	2015	Total
2009 cohort	13227.1	15164.9	11779.8	11895.5	10681.1	9307.4	9163.2	81219.1
2010 cohort	-	13478.7	14319.7	12802.0	11540.2	11001.2	10198.9	73340.7
2011 cohort	-	-	11301.5	14347.8	12298.4	10285.0	10366.7	58599.5
2012 cohort	-	-	-	9902.1	12979.9	11264.3	9807.7	43954.0
2013 cohort	-	-	-	-	9192.4	12503.5	11095.2	32791.1
2014 cohort	-	-	-	-	-	8519.7	12645.5	21165.3
2015 cohort	-	-	-	-	-	-	17338.0	17338.0
Total								328407.6

Table 9 shows how decrease in churn rate among the first-time visitors can contribute to the symphony center's ticket revenue. For each consumer cohort defined by which year they make first visit, I calculate how much incremental ticket revenues can be generated each year with 1.8% decrease in the churn rate after the first visit, assuming that the churn rate at each visit beyond the first visit stays intact.[12] On average, each cohort gives about $10K of incremental ticket revenue per year when the churn rate after the initial visit goes down by 1.8%. This results in $328K of cumulative incremental ticket revenue across different cohorts over the data period.

The significant impact of a single concert visit on customer churn at the symphony center level can be explained by consumers' incomplete information about the range of concert values offered by the symphony center. Having imperfect information about how representative the previously consumed concert is of what the symphony center offers, consumers who are exposed to low-value concerts in their early visits may generalize the information to all the other untried concerts and not come back, although the concerts they consumed lie in the low tail of the match value distribution. In sum, 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.

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 and Table 6), and 2) the correlation between the experienced concert values and customer returns in Table 8. The findings suggest that consumers who are exposed to low-value concerts (due to imperfect information at the purchase stage) simply abandon the symphony center afterwards. This pattern may arise because they generalize the experience to the center's other concerts they have not tried.

5.4 The effect of experienced concert value on customer lifetime value

Satisfying initial experience at the symphony center may not only change the return rate after the first visit but also affect other factors that affect customer lifetime value, including the length of tenure or the amount of donation. To see if there is any additional effect of initial concert experience on customer lifetime value, I run the same regression as in Equation (3) using two alternative outcome variables: conversion to a 'regular' consumer (with 10+ visits) and conversion to a donor. Table 10 summarizes the result.

[12] The table is calculated using the following equation:

ΔTicket Revenue in Year t from consumer cohort $c = 1.8\% \times$ (Observed ticket revenue generated by c in Year t).

Table 10: 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.

Coefficients on the experienced concert value ($\widehat{Q}_{i\tau}$) are significant and positive when the outcome variable is whether a consumer stays for more than 10 visits (Column 1 to 3). As in Table 8, 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 a conversion to a donor (Column 4 to 6). However, the correlation becomes insignificant for consumers who have stayed for three visits.

Table 10 also reports the predicted changes in the outcome variables under 1-standard-deviation increase in most recently experienced concert value ($\widehat{Q}_{i\tau}$).[13] When the average rate of conversion to a regular consumer among the first-time visitors is used, 1-standard-deviation increase in initially experienced concert value predicts about $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 9).

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.

5.5 The effect of experienced concert value on subsequent concert choice

Given the imperfect information that the new consumers have, experiences that new consumers get at initial concert visits can serve as a guideline for them to choose their next concert if they decide to return. In the context of search, Hodgson and Lewis (2018) show that a 'surprisingly bad' search result can lead consumers to choose a farther point for their next search because of the information spillover to nearby products in product feature space. Similarly, in the context of consumption, good or bad quality consumed in the current period can have an informative effect on what to consume next. For example, suppose a consumer experiences high quality from a concert. On one hand, she has more incentives to explore concerts that are different from the previously visited because now she has higher expectation about the overall concerts offered by the symphony

[13] To monetize contribution that changes in initial concert visits can make, I use the following equation to compute the predicted increase in ticket revenue and donation from 2009 consumer cohort for 7 years:

$$\begin{aligned}
\Delta\text{Ticket Revenue} = {} & \Delta(\text{Probablity of being a regular consumer}) \\
& \times (\text{Average total ticket price paid among regular consumers who make first visits in 2009} \\
& \quad \text{- Average total ticket price paid among non-regular consumers who make first visits in 2009}) \\
& \times \#\ \text{consumers who make first visits in 2009} \\
\Delta\text{Donation} = {} & \Delta(\text{Probablity of being a donor} \\
& \times (\text{Average total donation made consumers who make first visits in 2009}) \\
& \times \#\ \text{consumers who make first visits in 2009.}
\end{aligned}$$

Using median total ticket price or donation gives about 60% of the predicted values reported in Table 10.

center. On the other hand, she might be more willing to exploit the information she has gathered already by choosing the next concert to be similar to the previously visited.

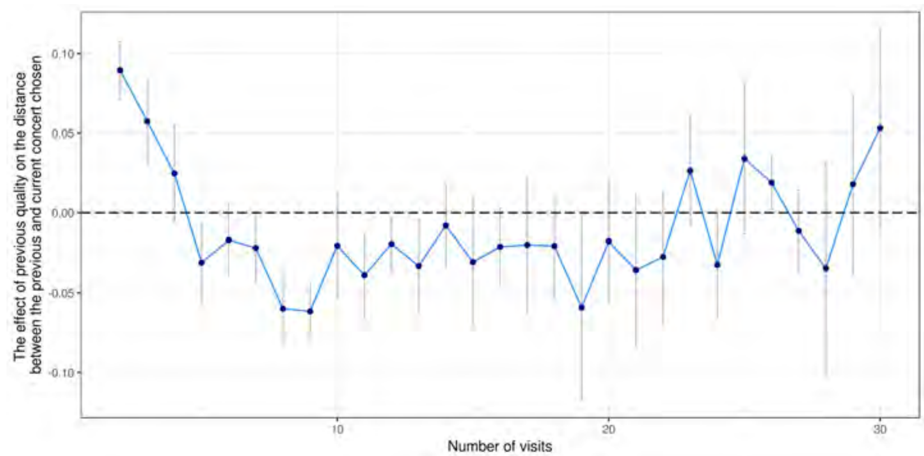
To see how people reflect recent concert experience in their subsequent product feature choice, I run the following regression:

$$d(X_{i\tau-1}, X_{i\tau}) = \sum_{t=2}^{30} \delta_t \cdot \mathbb{1}\{\tau = t\}\widehat{Q}_{i\tau-1} + Z_i\Gamma + \epsilon_{i\tau} \tag{5}$$

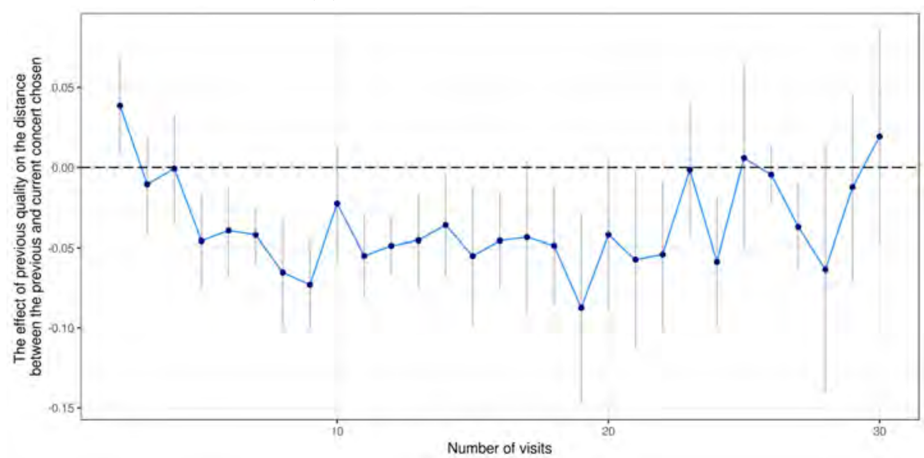
where $d(X_{i\tau-1}, X_{i\tau})$ is the distance in product feature space between the concert that i consumed in the $\tau - 1$-th visit and in the τ-th visit, $\widehat{Q}_{i\tau-1}$ is the concert value i consumed at her $\tau - 1$-th visit, and Z_i is the vector of control variables. δ_t denotes the effect of previous concert value on the product distance between the previous and the current choice, and this effect is modeled to be specific to the number of visits ($t \in \{2, \ldots, 50\}$). $d(\cdot, \cdot)$ is measured in several different ways using more than 600 binary and discrete variables and is normalized to have mean 0 and standard deviation of 1 in this particular regression. Detailed information on how the distance measures are calculated can be found in the Appendix.

Figure 8 reports the visit-specific coefficients of the previous concert value, δ_t. The x-axis denotes the number of visits, and the y-axis denotes the effect of previously experienced concert value on the distance (defined in the product feature space) between the previous and current concert choice. Three patterns are observed. First, coefficients of the previous concert value (δ_t) in the initial few visits are positive, implying that consumers explore more (i.e., purchase a product farther from the previous choice) at the current visit when the realized concert value in the previous visit is high. Second, the coefficients become negative after the first few visits. This indicates that, after certain number of visits, consumers choose similar products to the previously consumed one if their previous choice was of high quality. Third, the effect of the previous quality on the subsequent product choice becomes insignificant as the number of visits goes up. I find the same patterns using different distance metrics (See Section A.5 in the Appendix).

In summary, descriptive analyses provide the evidence that imperfect information causes customer attrition. In particular, 1) retention rate after a single visit is less than 50%, 2) new consumers have imperfect information on product quality variation at the purchase stage, and 3) the quality chosen in the current visit affects both return decision and product choice decision if return. Another piece of important descriptive evidence, which is discussed in Section 4, is that the average quality chosen by consumers increases as they make more visits (Figure 6, **??** in Section 4; Figure **??** in the Appendix).



(a) Using the entire sample



(b) Using consumers with 15+ visits

Figure 8: The effect of current quality on the next choice: the distance between the current and the next concert in the product space
(Using the distance measures caculated via logistic PCA)

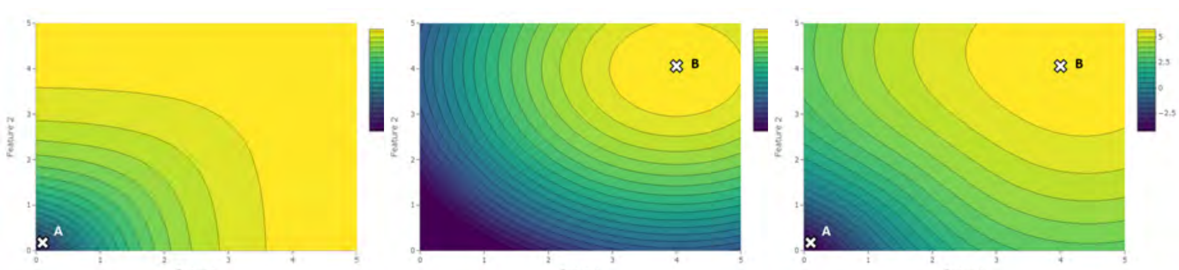
5.6 Discussion

Despite the richness of the existing learning models, most traditional Bayesian learning models cannot fully rationalize some of the 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 its potential limitations. The main points can be summarized as follows: 1) 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 qualities, 2) if consumers are myopic, the brand (symphony center)-level learning model should allow for very high signal variance to justify the high churn rate after a single visit (instead of normalizing it to 1 as in many previous studies), 3) the 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 quality products even among the products with new (untried) features.

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

6 A framework of consumer learning

I propose a new structural model of consumer learning that incorporates two types of information frictions, one at the purchase stage and the other at the updating stage. The model allows for flexible patterns of learning spillover 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 product of interest.



(a) Updating from the first purchase **(b)** Updating from the second purchase **(c)** Resulting belief after the two purchases

Figure 9: Illustration of learning spillover from consumption

Figure 9 shows an illustrative example of how learning spillover from consumption experience takes place in the product feature space. Two axes denote two product features available, and the color shade denotes the predicted value for each combination of the two product features. The brighter the color is, the higher the predicted value is. Initially, a consumer tries product A and experiences 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 9(a)). In the next purchase, she tries product B which is predicted to have high value based on her previous experience, which indeed gives positive experience (Figure 9(b)). The resulting belief after the two purchases about all the other untried products is the weighted average of the predicted qualities in (a) and (b) (Figure 9(c)).

The model captures several other data patterns that are not fully captured by traditional Bayesian learning models as discussed in Section 6.5. For example, attribute-level consumption-based learning models do not rationalize that consumers become better at selecting high-value products even when they face new product features they have not tried yet. To reflect this data pattern, I also allow consumers to obtain information through an additional channel besides consumption, which I call search. In this model, consumer decision to search is determined by how pleasant the previous product experiences were, which creates an incremental impact of prior product experiences on subsequent purchase behavior.

Consumers receive signals from both consumption and search activities. A signal from consumption is extrapolated to other untried products based on their similarity with the product consumed as illustrated in Figure 9. The intensity of a signal from search (i.e., how much a consumer engages in search to obtain additional information about untried products) is determined by how pleasant the past consumption experiences. 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. [14]

6.1 Utility

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

$$
\begin{aligned}
u_{ijt} &= U(Q_j^*, \epsilon_{ijt}) \\
u_{i0t} &= U(u_0, \epsilon_{i0t}) \qquad (6)
\end{aligned}
$$

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

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, consumer i makes purchase decisions based on the predicted consumption utility:

$$\begin{aligned} \widetilde{u}_{ijt} &= U(\widetilde{Q}_{it}(X_j), \epsilon_{ijt}) \\ \widetilde{u}_{i0t} &= U(u_0, \epsilon_{i0t}) \end{aligned} \tag{7}$$

where $\widetilde{u}$ denotes the predicted consumption utility and $\widetilde{Q}_{it}(X_j)$ denotes the product value predicted by consumer i at time t given the observable product characteristics X_j.[15]

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 product j among the available alternatives at time t ($\mathcal{J}_t = \{0, 1, \ldots, J_t\}$) as

$$Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\widetilde{v}_{ijt})}{\sum_{k \in \mathcal{J}_t} \exp(\widetilde{v}_{ikt})}. \tag{8}$$

6.2 Updating beliefs

Updating beliefs about product values through consumption After realizing the true value of product j, consumer i uses the difference between the true value and the predicted value to update her belief about all the untried products. Let $B_{it}(X_j)$ denote the prior belief on product j's value given the observable product features X_j, which is carried over from previous 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 and the belief on product value given previous consumption experiences:

$$\Delta_{ijt} = Q_j^* - B_{it}(X_j). \tag{9}$$

Then, consumer i's belief on 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 $d_{jk} = d(X_j, X_k)$):

$$B_{it+1}(X_k) = B_{it}(X_k) + g(N_{it}, d_{jk})\Delta_{ijt} \tag{10}$$

$$= B_{it}(X_k) + g_{ijkt}(Q_j^* - B_{it}(X_j)). \tag{11}$$

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

$g_{ijkt} = g(N_{it}, d_{jk})$ determines how much consumer i generalizes the information from a single experience Δ_{ijt} to other untried products based on product similarities. If $\frac{\partial g}{\partial d} = 0$, then the discrepancy Δ_{ijt} is equally applied to the belief of all the untried products regardless of the distance between the products and the previously consumed product (j in this case) in feature space. If $\frac{\partial g}{\partial d} < 0$, information spillover from consumption is discounted based on the similarity between the consumed product j and the untried product of interest. I assume that $|g(\cdot)| \leq 1$, i.e., the weight on the information spillover cannot exceed 1.

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

$$
\begin{aligned}
\widetilde{Q}_{it}(X_k) &= B_{it}(X_k) + S_{it}(X_k) \\
&= B_{it}(X_k) + \phi_{ikt}(Q_k^* - B_{it}(X_k)) && (12) \\
&= \phi_{it}Q_k^* + (1 - \phi_{it})B_{it}(X_k) && (13) \\
\text{where } \phi_{it} &= \text{information gain through search (or search intensity)} \\
&= \phi(\text{expected consumption benefit, expected search benefit,} && (14) \\
&\qquad \text{the number of past visits}). && (15)
\end{aligned}
$$

I assume that the number of visits affects search intensity because the quantity of advertising materials from the symphony center increases with the number of visits i has made in the past.[16] I assume that $\phi_{it} \in [0, 1]$; $\phi_{it} = 0$ means that there is no further correction in 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 to make a product choice.

Note from Equation (13) 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.

[16] This is because consumers get pamphlets and catalogs each time when they visit the symphony center, and because the symphony center send more informative materials about the upcoming concerts to consumers with more visits.

7 Model

I construct a structural model based on the framework in Section 7 and the assumptions motivated by the descriptive findings.

Consumers make purchase decisions every week given the product offerings of the week and an outside option (i.e., not purchasing any concert). Purchase decision is a function of predicted concert values. When purchase takes place, consumption utility is realized which is a function of an individual's preference over different genres and the true value of the concert within the specific genre. After she consumes a concert, the information she receives from consumption 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 consumed. The past consumption experiences also determine search intensity in the upcoming periods by forming consumers' expected consumption and search benefit.

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

$$\begin{aligned} u_{ijt} &= U(Q_j^*, x_{jt}, p_{jt}, \epsilon_{ijt}) = \gamma_{ig} + \eta_{it} + \delta_i \widehat{Q}_j^* - \alpha_i p_{jt} + \epsilon_{ijt} \\ u_{i0t} &= U(u_0, \epsilon_{i0t}) = u_0 + \epsilon_{i0t} \end{aligned} \tag{16}$$

where γ_{ig} is i's preference for genre g that concert j belongs to, η_{it} is the month, day, and hour fixed effect, and p_{jt} is the price of concert j. δ_i captures i's sensitivity to the experienced concert value $\widehat{Q}_j$. ϵ_{ijt} is a 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 the concert values, consumer i makes purchase decisions based on the predicted concert values:

$$\begin{aligned} \widetilde{u}_{ijt} &= \gamma_{ig} + \eta_{it} + \delta_i \widetilde{Q}_{it}(X_j) - \alpha_i p_{jt} + \epsilon_{ijt} \\ \widetilde{u}_{i0t} &= u_0 + \epsilon_{i0t} \end{aligned} \tag{17}$$

where $\widetilde{u}$ denotes the predicted consumption utility and $\widetilde{Q}_{it}(X_j)$ denotes the concert value predicted by consumer i at time t given the observable product characteristics X_j.[17]

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

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

of product j among the available alternatives at time t ($\mathcal{J}_t = \{0, 1, \ldots, J_t\}$) as

$$Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\widetilde{v}_{ijt})}{\sum_{k \in \mathcal{J}_t} \exp(\widetilde{v}_{ikt})}. \tag{18}$$

Updating belief about concert values through consumption After consuming concert j, consumer i uses the difference between the true value and the predicted value to update her belief about the upcoming concerts. Let Δ_{ijt} denote the discrepancy between the true and the predicted value based on the previous consumption experiences (as in Equation 9) :

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

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

$$B_{it+1}(X_k) = B_{it}(X_k) + g_{ijkt} \cdot \Delta_{ijt} \tag{19}$$

$$\text{where } g_{ijkt} = (\exp(\rho_{it} \cdot d(X_j, X_k)))^{-1}. \tag{20}$$

g_{ijkt} determines how much consumer i generalizes the information from a single consumption signal Δ_{ijt}; if $g_{ijkt} = 0$, then the discrepancy Δ_{ijt} is applied to all the upcoming concerts equally regardless of how far those upcoming concerts are in product feature space. I assume that g_{ijkt} is bounded above by 0. I model ρ_{it} to be the following:

$$\rho_{it} = \exp\left(\rho_{i0} + \rho_{i1} N_{it}\right). \tag{21}$$

Based on this specification, how local or global the experience spillover is varies across different levels of experiences. If $\rho_{i1} > 0$, it means that the spillover from previous experience takes place more locally as consumers make more visits, i.e., any concert experience only affects consumers' beliefs about similar concerts to the experienced one as consumers try more concerts. If $\rho_{i1} < 0$, this means that the spillover from concert experiences takes place more globally as consumers make more visits.

Updating belief about concert values through search Given the updated belief carried over from consumption experiences, consumers adjust their prediction on quality via search.

$$
\begin{aligned}
\widetilde{Q}_{it}(X_k) &= B_{it}(X_k) + \phi_{it}(Q_k^* - B_{it}(X_k)) \\
\text{where } \phi_{it} &= \text{information gain on true quality through search} \\
&= \phi(\text{expected consumption benefit, expected gain from search}) \\
&= \phi\left(\frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}} Q_j^*, \quad \frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}}\left[Q_j^* - \frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}} Q_j^*\right]^2\right) \\
\text{and } \phi_{it} &\in [0,1]
\end{aligned} \tag{22}
$$

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 upto week t. $\phi_{it} = 0$ means that there is no further correction in predicted concert value via search, while $\phi_{ijt} = 1$ implies that the true value is recovered via active search. I specify ϕ_{it} as follows:

$$
\begin{aligned}
\phi_{it} &= \frac{\phi_{i0} + \exp(\phi_{i1}a_{it} + \phi_{i2}b_{it} + \phi_{i3}N_{it})}{1 + \exp(\phi_{i0} + \phi_{i1}a_{it} + \phi_{i2}b_{it} + \phi_{i3}N_{it})} \\
\text{where } a_{it} &= \frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}} \widehat{Q}_j^* \\
\text{and } b_{it} &= \frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}}\left(Q_j^* - \frac{1}{N_{it}}\sum_{j\in\mathcal{N}_{it}} Q_j^*\right)^2.
\end{aligned} \tag{23}
$$

N_{it}, the number of visits, is included in the search intensity parameter because the quantity of advertising materials consumer i receives depends on how many visits she has made so far.

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

$$
\begin{aligned}
&\widetilde{Q}_{i0}(X_j) = \phi_{ip} Q_j^* \\
&\text{where} \quad \phi_{ip} = \frac{\exp(\phi_{i4})}{1 + \exp(\phi_{i4})}.
\end{aligned} \tag{24}
$$

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

Identification Clear identification strategy is required to decompose learning from search and learning from consumption when only purchase data is available (i.e., when there is no data on search). According to the model, beliefs about the concerts similar to the previously visited concerts are more affected by learning from consumption than beliefs about the concerts very different from the previously visited. Therefore, it does 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 the identification of the parameters that governs 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 adjusts her value prediction via search. Second, which *product features* she chooses when she returns helps identify how local or global the belief updating process is(ρ_{it}). If the experience was pleasant and she chooses the next concert to be very similar to the previous one, then the information from the previous consumption 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 in her subsequent visit, then it implies that the prior pleasant experience has globally updated her beliefs about all the untried concerts.

8 Estimation

For the 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.[18] Two different measures of product distance are constructed via logistic PCA and weighted Gower distance (See Appendix) and is scaled so that the maximum product distance is 1.

θ_i denotes a vector of utility parameters that are estimated:

$$\theta_i = \{\{\gamma_{i1}, \ldots, \gamma_{i8}\}, \{\eta_{i1}, \ldots, \eta_{i3}\}, \alpha_i, \delta_i, \{\rho_{i0}, \rho_{i1}\}, \{\phi_{i0}, \ldots, \phi_{i3}, \phi_{ip}\}\}.$$

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

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

$$
\begin{aligned}
\theta_i &= Z_i \Gamma_{s_i} + \iota_i \\
\iota_i &\sim N(0, \Sigma_{s_i}) \\
s_i &\sim \text{Multinomial}_K(\pi)
\end{aligned} \tag{25}
$$

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

Priors are defined as follows:

$$
\begin{aligned}
\Gamma &\sim N(\overline{\Gamma}_{s_i}, A^{-1}_{\Gamma_{s_i}}) \\
\pi &\sim \text{Dirichlet}(\overline{\alpha}) \\
\Sigma_s &\sim IW(\nu, V).
\end{aligned} \tag{26}
$$

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:

$$
\theta_i | s_i, Z_i\Gamma_{s_i}, \Sigma_{s_i} \tag{27}
$$

$$
\pi, s, \{\Gamma_s\}, \{\Sigma_s\} | \{\theta\}. \tag{28}
$$

I use Gaussian random-walk Metropolis step to draw θ_i given other sets of parameters (Step (27)). To make draws of mixture components (Step (28)), I use `rmultireg` and `rmixGibbs` from `bayesm` package (Rossi et al. (2005)).

9 Estimation Results

Table 11 and Figure 10 report the parameter estimates from the 2 mixture-of-normal model.[19] There is a major segment (Segment 1; 81.1% of the sample customers) and a minor segment (Segment 2; 18.9%) that are different in the tastes for specific genres, responses to the experienced concert values, and the amount of prior information about the underlying concert values.

First, I look at the estimated prior knowledge of customers by different segments. Figure 11 translates the prior knowledge parameter (ϕ_{i4}) to the fraction of the underlying concert values ($\phi_{ip} = \frac{exp(\phi_{i4})}{1+exp(\phi_{i4})}$; Equation 24). Although Segment 2 is reported to have more prior knowledge about the underlying concert values with a long right tail, both segments on average have little

[19] The estimation results from no-heterogeneity is in the Appendix.

Table 11: Distribution of posterior means of household coefficients (θ_i): 2 normal mixture components

	Min	1Q	Median	Mean	3Q	Max
Concert feature: Time						
Weekend	-8.08	-0.31	0.36	0.30	0.98	7.38
Summer	-5.46	-1.20	-0.39	-0.37	0.43	5.84
Evening	-8.41	-0.01	0.66	0.53	1.27	8.80
Concert feature: Genre						
Others	-18.89	-7.12	-5.98	-6.05	-4.89	5.10
Casual	-17.31	-2.73	-1.63	-1.54	-0.49	12.29
Chamber	-19.81	-3.21	-1.50	-1.60	0.16	13.59
Orchestra	-14.77	-1.97	-0.70	-0.65	0.58	13.41
Family	-20.24	-2.67	-1.26	-1.19	0.18	17.92
Jazz	-17.13	-1.70	-0.63	-0.53	0.48	14.33
Emerging professionals	-15.81	-2.87	-1.78	-1.94	-0.74	11.12
Specials	-10.76	-1.06	0.05	-0.13	1.04	6.13
Concert feature: Genre						
log(Price)	-441.10	-0.07	-0.04	-0.22	-0.02	-0.00
Predicted concert value	0.00	0.26	0.71	1.85	1.90	224.80
Experience spillover						
ρ_1	-5.13	0.08	0.91	0.95	1.76	8.80
ρ_2	-5.44	-0.58	0.19	0.20	0.96	6.62
Search intensity						
ϕ_0	-6.83	-0.85	0.00	0.02	0.87	8.11
ϕ_1	-7.03	-1.33	-0.32	-0.30	0.70	7.93
ϕ_2	-6.85	-0.33	0.76	0.80	1.89	8.26
ϕ_3	-12.03	-6.05	-5.00	-4.89	-3.89	3.55
Prior knowledge						
ϕ_4	-10.80	-5.77	-4.77	-4.75	-3.76	1.47

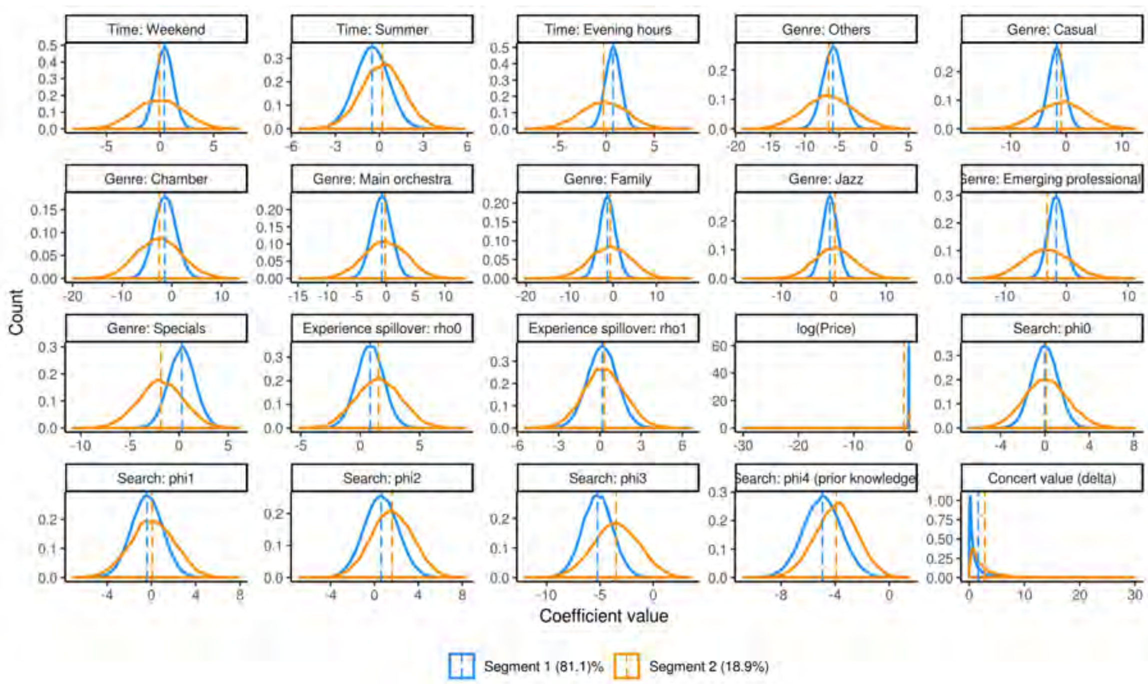


Figure 10: Distribution of posterior means of household coefficients (θ_i): 2 normal mixture components

information about the overall concert values.

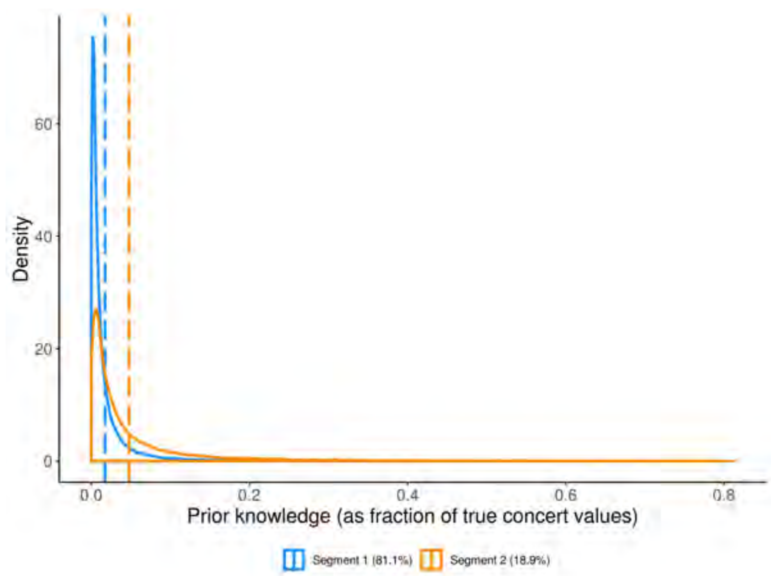
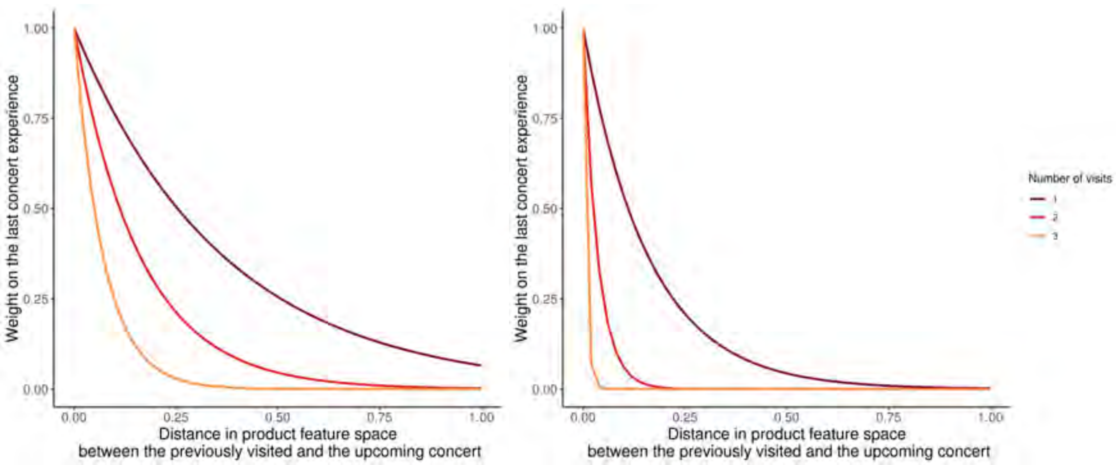


Figure 11: Distribution of prior information as a fraction of the actual concert values: 2 normal mixture components

Figure 12 illustrates the degrees of information spillover at each visit given the parameter estimates. The x-axis denotes the distance in feature space from the previously visited concert to any upcoming concert under evaluation, and the y-axis is the degree of experience spillover from the previous concert experience. For example, $x = 0.5$ and $y = 0.3$ indicates that the predicted concert value of any upcoming concerts that are located 0.5 away from the previously visited concert j is updated by $0.3 \times \Delta_{ij}$(information discrepancy for concert j). Each cell represents predicted spillover pattern of customers in each segment (normal mixture component). Two patterns are noticeable. First, the experienced concert value at the first visit affects the prediction of the concert values in the entire product space, especially for Segment 1 (dark red line). This suggests that the experience at the first visit is generalized to the entire offerings (i.e., the information update happens in a global manner) no matter how different or similar the upcoming concerts are to the initially visited. Second, the spillover takes place more locally as the number of visits goes up (light orange line). According to these results, the same set of concerts can generate different perceptions of the entire concert value distribution if the concerts are experienced in different orders.



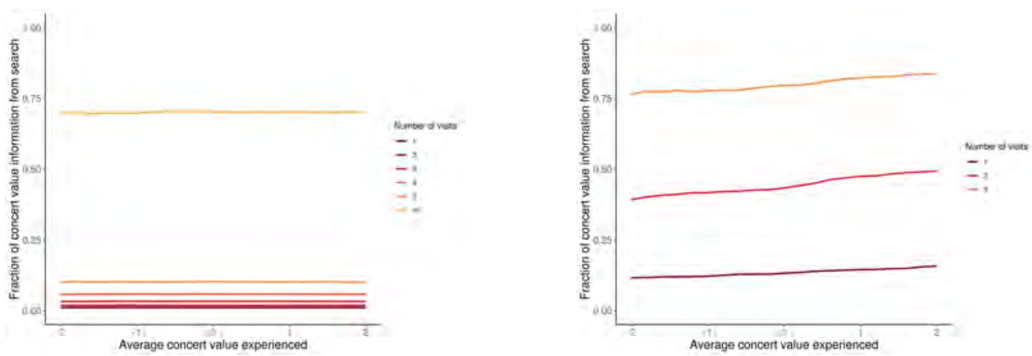
(a) Segment 1 (81.1%)

(b) Segment 2 (18.9%)

Figure 12: Experience spillover: 2 normal mixture components (using median of the distribution of posterior means of household coefficients)

Figure 13 shows how the search intensity $\phi_{it} \in [0, 1]$ changes over visits and with the average concert value experienced. The x-axis denotes the concert value at the initial visit, and the y-axis denotes the intensity of the signal from search as a fraction of the true concert values. Color-coded lines refer to different numbers of visits. The graphs imply that the intensity of search, or the degree of adjustment of concert value prediction formed from the previous experiences, varies significantly across the number of visits. For Segment 1, the degree of search is minimal after the first few visits, and it becomes significant as they make many more visits (10+). For Segment 2, the degree of search is also not very high after the first visit although much higher than Segment 1's. However, it increases much faster over the visits so that the search intensity becomes greater than 0.75 after 3 visits. After taking into account the effect of number of visits on search intensity, I find that the effect of average experienced concert values on search intensity to be much less significant.

Low prior information about the underlying concert values and the two patterns of information acquisition jointly justify the early attrition at the symphony center level as well as offer managerial insights on how to prevent the attrition. Given the lack of prior information about the concert values at the purchase stage, new consumers who randomly purchase low-value concerts in their first visits would only generalize the information (Figure 12) without adjusting it via search (Figure 13), which results in their churn at the symphony center level. Both information acquisition patterns highlight that the order of the concert consumption matters in forming people's perception about the entire concert offerings; initial concert experiences are used to update people's perception about other upcoming concerts more globally, and the resulting perception is less adjusted by additional search during the early visits. This suggests that treating new customers with high-value concerts during the initial visits may have a lasting impact on customers' perception.



(a) Segment 1 (81.1%) **(b)** Segment 2 (18.9%)

Figure 13: Experience spillover: 2 normal mixture components (using median of the distribution of posterior means of household coefficients)

10 Counterfactuals

Two sets of counterfactuals are run with different goals. In the first set, the goal is to revert churn, i.e., to re-attract consumers who have already exposed to the low-quality concerts in their first visits. In the second set, the goal is to prevent churn, i.e., to raise the average consumer tenure across all consumers.

10.1 Scenario 1: Reverting churn

In this scenario, a control group consists of consumers who have low-quality concerts in their first visits. The objective is to re-attract these consumers, i.e., raising the average number of visits in this specific group of consumers via marketing interventions.

Given the structural parameter estimates, I simulate the control group by taking the actual consumer data (zip code, the week of the first purchase, etc.) and assigning each consumer to the lowest-quality concert of the week for their initial purchase. Three types of marketing interventions are applied to this control group: 1) Recommendation (correcting the first concert to the high-quality concert), 2) Increasing search intensity after the first visit (e.g., sending more direct mails, catalogs, etc.), and 3) Price promotion for the second visits. The first intervention (Recommendation) is assumed to take place before the first visit to the low-quality concerts, and the other two interventions are assumed to happen after the first visit. Each type has two different levels. For Search treatment, I test 0.1 increase and 0.5 increase in search intensity. For Price treatment, I test 10% discount and 50% discount. For Recommendation treatment, I test two different strategies. The first strategy is to assign each consumer to the highest-quality concert of the week. The second strategy is to make the same recommendation (assignment) and additionally to remove low quality concerts from the offering. In the second strategy, any concert that has quality lower than or equal to -0.5 is treated as unavailable, which is the low 7% of the entire product offerings. This is equivalent to dropping 14 concerts per year.

Under different interventions, purchase sequence for each consumer is simulated and the total number of visits is averaged across the consumers over 100 weeks.

Table 12 reports the counterfactual results. Under Recommendation strategy, the average number of return visits over 100 weeks increases from 2.19 to 2.43, which is 10% increase in the average return visit and 7.5% increase in the average total visit (including the first visit). The increase is significant even if the low-quality products are completely dropped, although the size of the increase is lower than in the case of recommendation only (6.2% increase in the average return visit and 4.3% increase in the average total visit). This hints the trade-off between increasing the number of the product offerings and increasing the average quality of the product offerings; the baseline probability that a purchase takes place is higher when the symphony center provides more products, but having more products also negatively affects the customer lifetime value if the chance is higher that consumers purchase low-quality concerts in their first visits.

Increasing post-purchase search intensity by 0.1 via marketing treatments (e.g., sending more catalogs) does not result in significant increase in the average number of visits (Line 4 of Table 12). Increasing search intensity[20] by 0.5 raises the average number of visits by 1.9%, which is much less than the percentage increase under Recommendation strategies. The results are similar under price promotions. 10% discount on the next purchase does not have a significant impact on consumers' return visits. Even having 50% discount for any concert on the second visit is not enough to match the effect of having high-quality concerts on the first visits.

In summary, the first set of counterfactual exercises highlights the importance of the first expe-

[20]Note that the search intensity parameter is bounded between 0 and 1 in my model: $\phi \in [0, 1]$.

Table 12: Counterfactual results - Average number of visits over 100 weeks

Counterfactual scenario	Av. number of return visits (excluding the first visit)	= Control mean? (p-value from the t-test)
Control (having low-quality concerts as the first visit)	2.1866	-
Recommendation	2.4253	< 0.00005
Recommendation + Dropping-low quality products	2.3241	< 0.00005
Increase search intensity (+0.1)	2.1995	0.4656
Increase search intensity (+0.5)	2.2438	0.0033
Price promotion for second visit (10%)	2.1954	0.6851
Price promotion for second visit (50%)	2.2258	0.0684

rience given the nature of learning; the firm should pay consumers a large amount of compensation to make them return if they are initially exposed to negative product experiences. This finding underlines the significance of product recommendation system and well-curated choice architecture especially for the potential customers with no past experience.

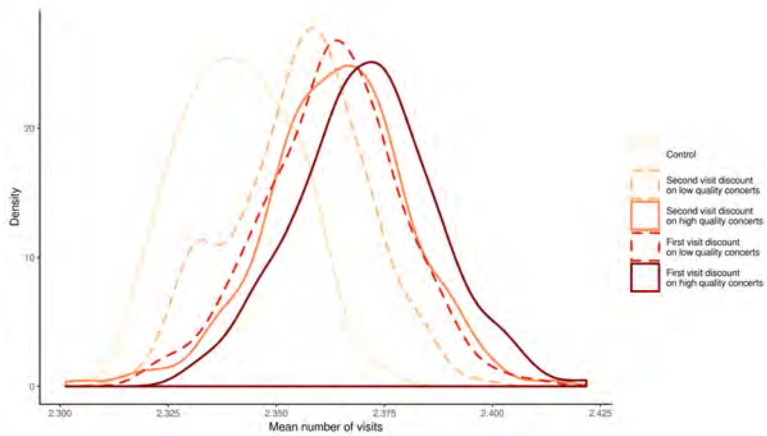
10.2 Scenario 2: Preventing churn

In this scenario, a control group consists of the entire sample of the consumers selected for structural estimation. Here, I allow consumers to make zero purchase during 100 weeks. The outcome measure of interest is the unconditional average number of visits (including 0 purchases) under different marketing interventions.

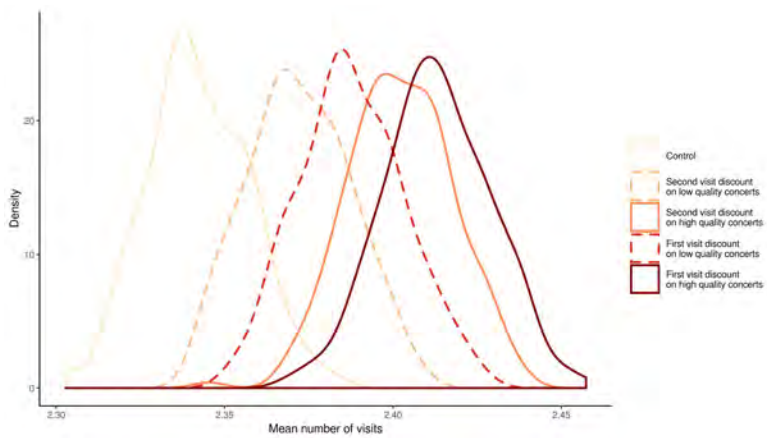
The marketing strategies tested here are designed to be 2 by 2: the time of price promotion (on the first visit vs. on the second visit) $\times$ the products on price promotion (low-quality products vs. high-quality products). Both dimensions are relevant to the real-world price promotion settings. As for the time dimension, we find both a sign-up discount for the first purchase and a discount coupon for the next purchase to be common. We also see either popular items and less popular items on discounts for different reasons (promoting new menus in the restaurant vs. having a clearance section to remove inventories). By testing the four different strategies relevant to the real-world setting, I compare the consequence that each strategy can have.

I simulate the purchase sequence for each consumer using the real consumer data and structural estimates (using Equation 16 to 23). I define high quality products to be any concert with $\widehat{Q}_j > 0$ and low quality products to be the ones with $\widehat{Q}_j < 0$. Unlike in the first scenario in which I force each consumer to choose the lowest quality concert in her Week 1, I allow zero purchase to happen here. To report the results, I bootstrap for 1,000 times and report the distribution of the average numbers of visits under different marketing interventions.

Figure 14 shows the bootstrap results when the price discount is set to be 50% and 90%. According to Figure 14(a), the unconditional average number of visits (including 0 visits) is the highest under the price promotion on high-quality products to the first-timers only. Promoting low-quality concerts results in lower average number of visits than promoting high-quality concerts does (dashed lines vs. solid lines). It implies that price promotions on low-quality products can have a negative effect on customer base, because it can nudge people to first consume low-quality concerts and update their brand perception accordingly. Figure 14(b) reports the same results at the discount rate of 90%, which highlights the differences among the four strategies even more.



(a) 50% discount



(b) 90% discount

Figure 14: Average number of visits under different price promotions (including 0 visits)

11 Conclusion

This paper shows how information frictions and consumer learning affect customer attrition among new customers in the context of classical music concerts. After documenting the evidence of imperfect information by testing for consumer learning, I descriptively show that two types of information frictions jointly impact consumer return decisions. Random arrival rates of new customers to concerts with varying qualities suggest that there exist information frictions on product qualities at the purchase stage. A significant impact of a single product experience on the subsequent churn rate indicates that consumers treat the single experience to be highly representative of what the symphony center offers. This implies that there exist incomplete information about the variance in available qualities at the learning stage. These two pieces of evidence imply the underlying mechanism behind why consumers abandon a brand after a single product trial: Consumers who buy low-quality products (due to the information frictions at the purchase stage) generalize the negative experience to all the other untried items and leave the brand (due to the information frictions at the learning stage).

The new framework proposed in this paper incorporates these information frictions and few other descriptive findings by allowing consumers to have flexible pattern of learning spillover and to obtain additional information through endogenous search behavior. Counterfactual analyses highlight that optimal marketing strategies should explicitly consider the existence of incomplete information at both purchase and updating stage. 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 consumer brand abandonment might be an important issue for a firm to solve.

The paper sheds light on the importance of initial experiences in consumer-brand relationships, which has been documented in consumer psychology literature but not in Bayesian consumer learning models. While different theories and heuristics imply the potential significance of the first information on subsequent updating, Bayesian learning models have assumed that any information contained in the initial signal would be debiased via following purchases. The gap exists partly because it is econometrically challenging to identify the causal impact of the realized initial experience from an unobservable random utility component from the purchase data only. I overcome this issue by using the quality measures inferred by experienced consumer choices.

In summary, this paper studies the underlying mechanism of consumer churn motivated by rich data patterns and consumer learning theories. More broadly, it opens a discussion on what a brand means in a modern context - a collection of diverse experiences instead of a small number of standardized products - and how consumers learn about the value of a brand based on few samples of experiences. Based on the general framework proposed in this paper, future research may extend our knowledge on the role of imperfect information and consumer learning on the optimal marketing strategies at the brand level.

References

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal* *of Marketing Research*, *55*(1), 80-98. Retrieved from https://doi.org/10.1509/jmr.16.0163 doi: 10.1509/jmr.16.0163

Ascarza, E., & Hardie, B. G. S. (2013). A joint model of usage and churn in contractual settings. *Marketing Science*, *32*(4), 570-590. Retrieved from https://doi.org/10.1287/mksc.2013.0786 doi: 10.1287/mksc.2013.0786

Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, *53*(1), 46-60. Retrieved from https://doi.org/10.1509/jmr.13.0483 doi: 10.1509/jmr.13.0483

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., ... Schrift, R. (2018, Mar 01). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, *5*(1), 65–81. Retrieved from https://doi.org/10.1007/s40547-017-0080-0 doi: 10.1007/s40547-017-0080-0

Ascarza, E., Netzer, O., & Hardie, B. G. S. (2018). Some customers would rather leave without saying goodbye. *Marketing Science*, *37*(1), 54-77. Retrieved from https://doi.org/10.1287/mksc.2017.1057 doi: 10.1287/mksc.2017.1057

Ater, I., & Landsman, V. (2013). Do customers learn from experience? evidence from retail banking. *Management Science*, *59*(9), 2019-2035. Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1120.1694 doi: 10.1287/mnsc.1120.1694

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, *25*(2), 242–262. Retrieved from http://www.jstor.org/stable/2555829

Blattberg, R. C., Kim, B.-D., & Neslin, S. A. (2008). *Database marketing* (1st ed., Vol. 18). Springer-Verlag New York.

Bolton, R. N. (1998). 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. Retrieved from https://doi.org/10.1287/mksc.17.1.45 doi: 10.1287/mksc.17.1.45

Bolton, R. N., Lemon, K. N., & Bramlett, M. D. (2006). The effect of service experiences over time on a supplier's retention of business customers. *Management Science*, *52*(12), 1811-1823. Retrieved from https://doi.org/10.1287/mnsc.1060.0594 doi: 10.1287/mnsc.1060.0594

Borle, S., Boatwright, P., Kadane, J. B., Nunes, J. C., & Galit, S. (2005). The effect of product assortment changes on customer retention. *Marketing Science*, *24*(4), 616-622. Retrieved from https://doi.org/10.1287/mksc.1050.0121 doi: 10.1287/mksc.1050.0121

Braun, M., & Schweidel, D. A. (2011). Modeling customer lifetimes with multiple causes of churn. *Marketing Science*, *30*(5), 881-902. Retrieved from https://doi.org/10.1287/mksc.1110.0665 doi: 10.1287/mksc.1110.0665

Braun, M., Schweidel, D. A., & Stein, E. (2015). Transaction attributes and customer valuation. *Journal of Marketing Research*, *52*(6), 848-864. Retrieved from https://doi.org/10.1509/jmr.13.0377 doi: 10.1509/jmr.13.0377

Caetano, G. (2016, August). *Neighborhood sorting and the value of public school quality.*

Camacho, N., Donkers, B., & Stremersch, S. (2011). Predictably non-bayesian: Quantifying salience effects in physician learning about drug quality. *Marketing Science*, *30*(2), 305-320. Retrieved from https://doi.org/10.1287/mksc.1100.0624 doi: 10.1287/mksc.1100.0624

Capraro, A. J., Broniarczyk, S., & Srivastava, R. K. (2003). Factors influencing the likelihood of customer defection: The role of consumer knowledge. *Journal of the Academy of Marketing Science*, *31*(2), 164-175. Retrieved from https://doi.org/10.1177/0092070302250900 doi: 10.1177/0092070302250900

Che, H., Erdem, T., & Öncü, T. S. (2015, Sep 01). Consumer learning and evolution of consumer brand preferences. *Quantitative Marketing and Economics*, *13*(3), 173–202. Retrieved from https://doi.org/10.1007/s11129-015-9158-x doi: 10.1007/s11129-015-9158-x

Ching, A., Erdem, T., & Keane, M. P. (2013). Invited paper—learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, *32*(6), 913-938. Retrieved from https://doi.org/10.1287/mksc.2013.0805 doi: 10.1287/mksc.2013.0805

Ching, A., & Lim, H. (2019). A structural model of correlated learning and late-mover advantages: The case of statins. *Management Science*, *Forthcoming*. doi: 10.2139/ssrn.2662286

Chintagunta, P. K., Goettler, R. L., & Kim, M. (2012). New drug diffusion when forward-looking physicians learn from patient feedback and detailing. *Journal of Marketing Research*, *49*(6), 807-821. Retrieved from https://doi.org/10.1509/jmr.11.0114 doi: 10.1509/jmr.11.0114

Chintagunta, P. K., Jiang, R., & Jin, G. Z. (2009, Dec 01). Information, learning, and drug diffusion: The case of cox-2 inhibitors. *QME*, *7*(4), 399–443. Retrieved from https://doi.org/10.1007/s11129-009-9072-1 doi: 10.1007/s11129-009-9072-1

Coscelli, A., & Shum, M. (2004). An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*, *122*(2), 213 - 246. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407603002744 doi: https://doi.org/10.1016/j.jeconom.2003.09.002

Dempster, A., Laird, N., & B. Rubin, D. (1977, 01). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*, 1-38.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523-538. Retrieved from https://doi.org/10.1287/orsc.12.5.523.10092 doi: 10.1287/orsc.12.5.523.10092

Dubé, J.-P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, *41*(3), 417-445. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-2171.2010.00106.x doi: 10.1111/j.1756-2171.2010.00106.x

Epstein, L. G. (2006, 04). An Axiomatic Model of Non-Bayesian Updating. *The Review of Economic Studies*, *73*(2), 413-436. Retrieved from https://doi.org/10.1111/j.1467-937X.2006.00381.x doi: 10.1111/j.1467-937X.2006.00381.x

Erdem, T. (1998). An empirical analysis of umbrella branding. *Journal of Marketing Research*, *35*(3), 339-351. Retrieved from https://doi.org/10.1177/002224379803500305 doi: 10.1177/002224379803500305

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, *15*(1), 1–20. Retrieved from `http://www.jstor.org/stable/184181`

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005b). Rfm and clv: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, *42*(4), 415-430. Retrieved from `https://doi.org/10.1509/jmkr.2005.42.4.415` doi: 10.1509/jmkr.2005.42.4.415

Fader, P. S., & Hardie, B. G. S. (2010). Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science*, *29*(1), 85-93. Retrieved from `https://doi.org/10.1287/mksc.1080.0482` doi: 10.1287/mksc.1080.0482

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005a). "counting your customers" the easy way: An alternative to the pareto/nbd model. *Marketing Science*, *24*(2), 275-284. Retrieved from `https://doi.org/10.1287/mksc.1040.0098` doi: 10.1287/mksc.1040.0098

Fader, P. S., Hardie, B. G. S., & Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, *29*(6), 1086-1108. Retrieved from `https://doi.org/10.1287/mksc.1100.0580` doi: 10.1287/mksc.1100.0580

Fryer, J., Roland G, Harms, P., & Jackson, M. O. (2018, 08). Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization. *Journal of the European Economic Association*. Retrieved from `https://doi.org/10.1093/jeea/jvy025` doi: 10.1093/jeea/jvy025

Haggag, K., Pope, D. G., Bryant-Lees, K. B., & Bos, M. W. (2018, 09). Attribution Bias in Consumer Choice. *The Review of Economic Studies*. Retrieved from `https://doi.org/10.1093/restud/rdy054` doi: 10.1093/restud/rdy054

Heilman, C. M., Bowman, D., & Wright, G. P. (2000). The evolution of brand preferences and choice behaviors of consumers new to a market. *Journal of Marketing Research*, *37*(2), 139–155. Retrieved from `http://www.jstor.org/stable/1558496`

Hodgson, C., & Lewis, G. (2018, November). *You can lead a horse to water: Spatial learning and path dependence in consumer search.*

Israel, M. (2005, December). Services as experience goods: An empirical examination of consumer learning in automobile insurance. *American Economic Review*, *95*(5), 1444-1463. Retrieved from `http://www.aeaweb.org/articles?id=10.1257/000282805775014335` doi: 10.1257/000282805775014335

Iyengar, R., Ansari, A., & Gupta, S. (2007). A model of consumer learning for service quality and usage. *Journal of Marketing Research*, *44*(4), 529-544. Retrieved from `https://doi.org/10.1509/jmkr.44.4.529` doi: 10.1509/jmkr.44.4.529

Jaffe, S., Coles, P., Levitt, S., & Popov, I. (2019, March). *Quality externalities on platforms: The case of airbnb.*

Kardes, F. R. (1986, 06). Effects of Initial Product Judgments on Subsequent Memory-Based Judgments. *Journal of Consumer Research*, *13*(1), 1-11. Retrieved from `https://doi.org/10.1086/209043` doi: 10.1086/209043

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311–329. Retrieved from http://www.jstor.org/stable/1830691

Nosko, C., & Tadelis, S. (2015, February). *The limits of reputation in platform markets: An empirical analysis and field experiment.*

Orhun, A. Y., Venkataraman, S., & Chintagunta, P. K. (2016). Impact of competition on product decisions: Movie choices of exhibitors. *Marketing Science*, *35*(1), 73-92. Retrieved from https://doi.org/10.1287/mksc.2015.0909 doi: 10.1287/mksc.2015.0909

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*(5), 879. Retrieved from http://proxy.uchicago.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edscal&AN=edscal.15163090&site=eds-live&scope=site

Rabin, M., & Schrag, J. L. (1999, 02). First Impressions Matter: A Model of Confirmatory Bias*. *The Quarterly Journal of Economics*, *114*(1), 37-82. Retrieved from https://doi.org/10.1162/003355399555945 doi: 10.1162/003355399555945

Reichheld, F. F. (1996, March). *Learning from customer defections.* Harvard Business Review.

Robalo, P., & Sayag, R. (2018). Paying is believing: The effect of costly information on bayesian updating. *Journal of Economic Behavior & Organization*, *156*, 114 - 125. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167268118302646 doi: https://doi.org/10.1016/j.jebo.2018.09.016

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian statistics and marketing* (1st ed.). Wiley.

Rust, R. T., Inman, J. J., Jia, J., & Zahorik, A. (1999). What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science*, *18*(1), 77-92. Retrieved from https://doi.org/10.1287/mksc.18.1.77 doi: 10.1287/mksc.18.1.77

Sanbonmatsu, D. M., Kardes, F. R., Houghton, D. C., Ho, E. A., & Posavac, S. S. (2003). Overestimating the importance of the given information in multiattribute consumer judgment. *Journal of Consumer Psychology*, *13*(3), 289-300. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1207/S15327663JCP1303_10 doi: 10.1207/S15327663JCP1303_10

Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, *33*(1), 1–24. Retrieved from http://www.jstor.org/stable/2631608

Schweidel, D. A., Park, Y.-H., & Jamal, Z. (2014). A multiactivity latent attrition model for customer base analysis. *Marketing Science*, *33*(2), 273-286. Retrieved from https://doi.org/10.1287/mksc.2013.0832 doi: 10.1287/mksc.2013.0832

Shin, S., Misra, S., & Horsky, D. (2012). Disentangling preferences and learning in brand choice models. *Marketing Science*, *31*(1), 115-137. Retrieved from https://doi.org/10.1287/mksc.1110.0680 doi: 10.1287/mksc.1110.0680

Sridhar, K., Bezawada, R., & Trivedi, M. (2012). Investigating the drivers of consumer cross-category learning for new products using multiple data sets. *Marketing Science*, *31*(4), 668-688. Retrieved from `https://doi.org/10.1287/mksc.1120.0717` doi: 10.1287/mksc.1120.0717

Sriram, S., & Chintagunta, P. K. (2010, January). Learning models. In N. K. Malhotra (Ed.), (Vol. 6, pp. 63–83). Emerald Group Publishing Limited. Retrieved from `https://doi.org/10.1108/S1548-6435(2009)0000006007`

Sriram, S., Chintagunta, P. K., & Manchanda, P. (2015). Service quality variability and termination behavior. *Management Science*, *61*(11), 2739-2759. Retrieved from `https://doi.org/10.1287/mnsc.2014.2105` doi: 10.1287/mnsc.2014.2105

Stiglitz, J. E. (1989). Chapter 13 imperfect information in the product market. In (Vol. 1, p. 769 - 847). Elsevier. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1573448X89010162` doi: https://doi.org/10.1016/S1573-448X(89)01016-2

Szymanowski, M., & Gijsbrechts, E. (2012). Consumption-based cross-brand learning: Are private labels really private? *Journal of Marketing Research*, *49*(2), 231-246. Retrieved from `https://doi.org/10.1509/jmr.07.0416` doi: 10.1509/jmr.07.0416

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. Retrieved from `https://science.sciencemag.org/content/185/4157/1124` doi: 10.1126/science.185.4157.1124

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599-637. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12101` doi: 10.1111/cogs.12101

White, A., & Murphy, T. (2014). Bayeslca: An r package for bayesian latent class analysis. *Journal of Statistical Software, Articles*, *61*(13), 1–28. Retrieved from `https://www.jstatsoft.org/v061/i13` doi: 10.18637/jss.v061.i13

Wyer, R. S., Srull, T. K., & Gordon, S. (1984). The effects of predicting a person's behavior on subsequent trait judgments. *Journal of Experimental Social Psychology*, *20*(1), 29 - 46. Retrieved from `http://www.sciencedirect.com/science/article/pii/0022103184900106` doi: https://doi.org/10.1016/0022-1031(84)90010-6

Zhang, Y., Bradlow, E. T., & Small, D. S. (2015). Predicting customer value using clumpiness: From rfm to rfmc. *Marketing Science*, *34*(2), 195-208. Retrieved from `https://doi.org/10.1287/mksc.2014.0873` doi: 10.1287/mksc.2014.0873

Appendix

A.1 Specification of $\widehat{Q}_{jt}^{R}$: Heterogeneous preference over genres

The inclusion of genre fixed effects in Equation **??** and 2 allows that genre preferences are already known to consumers instead of being learned. In other words, by constructing a quality measure that controls for average genre preferences of the experienced consumers, I assume that there is hidden quality to be learned across concerts within a genre, but consumers have full information about their preferences for each 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, the choice of which product features are included in hidden qualities to be learned and which are included in known preferences is made by researchers. The reason why I categorize genre preference as known and all the other feature preferences as unknown is because the data shows that the first-timers' arrival rate to different concerts is uniform *within* any genre. Robustness checks with a quality measure without genre fixed effects confirms that the major descriptive patterns stay qualitatively the same but much more noisy. Future research may investigate how to determine the structure in a more informed way.

A.2 Documentation of learning

Figure 15 in the Appendix shows the average concert values chosen at each visit by consumer groups with different length of observed tenure. Most groups, regardless of how many total they have made during the data period, exhibit the upward sloping pattern between the number of visits and the average concert value purchased, which supports the model of consumer learning about the underlying concert values.

A.3 Validity of the estimated concert values $\widehat{Q}^{*}$

To further check the validity of the estimated quality measure, I scrape the names of top 20 classical music albums in Billboard's weekly chart during the data period and compare them with the estimated concert values.[21].

[21] https://www.billboard.com/charts/classical-albums

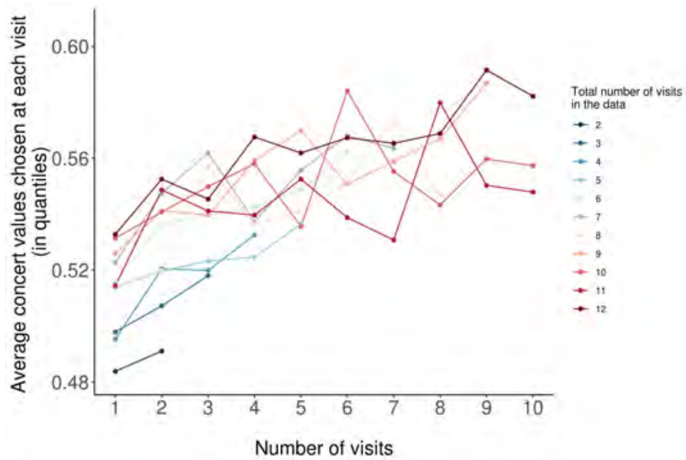


Figure 15: Average concert values chosen at each visit, by observed consumer tenure

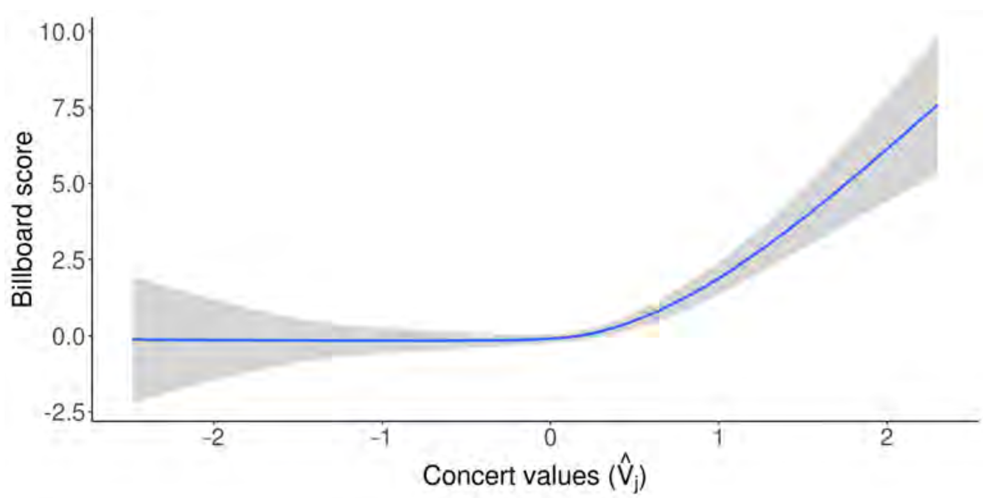


Figure 16: Correlation between the estimated concert values $\widehat{Q}_j^*$ and Billboard rankings

Next, I run the following regression:

$$\begin{aligned} Q_j^* \text{ chosen by consumer } i = &\text{Individual FE} \\ &+ \beta_1 \log(\#\text{Days between the performance and the ticket order date}) \\ &+ \beta_2 \log(\text{Price paid}) + \beta_3 \mathbb{1}\{\text{Purchased Seat Quantity} = 1\} + \epsilon_{ij}. \end{aligned} \tag{A.0.1}$$

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

Inclusion of any ticket bundle purchase may contaminate the result because the performance dates of the concerts included in any given bundle spread out widely across the season and therefore can introduce spurious variation in the days of wait. Therefore, I only include non-bundle-ticket purchases by experienced consumers with more than 15 past visits in this regression.

Table 13 shows that all coefficients are positive as predicted.

Table 13: Correlation between the 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)
$\mathbb{1}$\{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

One could argue that there might be systematic difference in tastes between the experienced consumers 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 14 shows the list of top ranked concerts according to the measure. The list confirms that the concert value measure is not biased against those concerts targeting the mass public.

Table 14: Top ranked concerts according to the quality 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.4 Heterogeneity in Q^*

The estimator can easily be modified to explicitly allow for heterogeneity in perceived product quality. One way to incorporate consumer-level heterogeneity is to cluster experienced consumers into latent classes based on their purchase sequences and to construct separate quality measures for each segment. Table 15 illustrates the approach.

Table 15: Clustering experienced consumers based on purchase decisions

Consumer ID	Concert 1	Concert 2	...	Concert $J-1$	Concert J
A	1	1	...	0	0
B	1	1	...	0	0
C	1	0	...	0	0
D	1	1	...	0	0
E	0	1	...	1	1
F	0	1	...	1	1
Total share	67%	83%	...	33%	33%
Segment 1 share	100%	75%	...	0	0
Segment 2 share	0	100%	...	100%	100%

Here, clustering six consumers into two segments - {A, B, C, D} and {E, F} - gives two sets of within-segment market shares different from the total shares. Although total choice shares of experienced consumers can be viewed as market average quality measures, recovering latent groups of expert consumers based on their choices offers richer information on heterogeneity in preferences over different concerts.

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

Let $Y = (Y_1, \ldots, Y_N)$ denote a binary vector of concert purchases by N experienced consumers where $Y_i = (y_{i1}, \ldots, y_{iJ})$ and J is the total number of concert offerings. Each consumer 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 lable 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 class-specific purchase probabilities and class 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 segments ($G = 2$) estimated via this algorithm are presented in Figure 17.

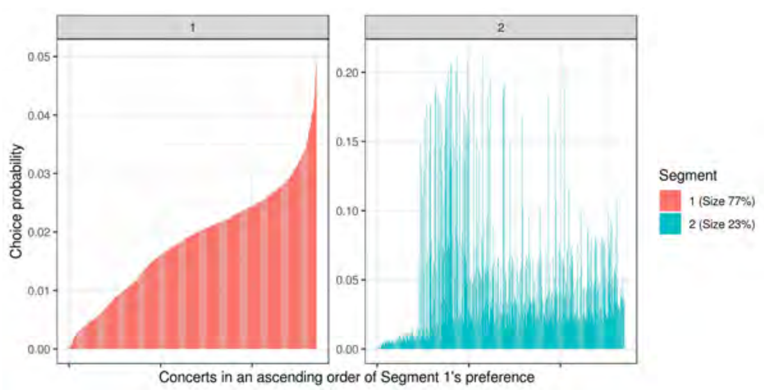


Figure 17: Estimated choice probabilities of experienced consumers by latent segment

Figure 17 shows the estimated choice probabilities of individual concerts assuming two different segments. The x-axis denotes individual concert ordered by Segment 1's preference, and the y-axis represents the choice probability of each concert by segment. Two patterns are noticeable. First, there is a big asymmetry in size between Segment 1 and Segment 2, which implies that the majority of consumers are clustered as the same latent class. Second, both segments 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.

A.5 The effect of consumption experiences on subsequent product choice: Using different distance metrics in product space

As a robustness check, I use different specifications of product distance and see how they affect the results. Two different approaches are used to construct a distance measure given a large set of binary variables:

- *Logistic Principal Component Analysis (PCA)* (Langradf and Lee 2015): I first run Logistic Principal Component Analysis on 615 binary variables of 1350 concerts setting the number of components to be 100 (whose result explains 98.1% of the data variance). Using the resulting 100 components that are continuous, I calculate the euclidean distance between concerts. I also vary the weights attached to each binary variable when doing the logistic PCA step (e.g., the total number of appearances across concerts as the weight for each dummy variable).
- *Gower's metric* (Gower 1971): Gower's metric is defined as $G_{ij} = \sum_k d_{ijk}$ where d_{ijk} is the difference between product i and concert j in product feature k.

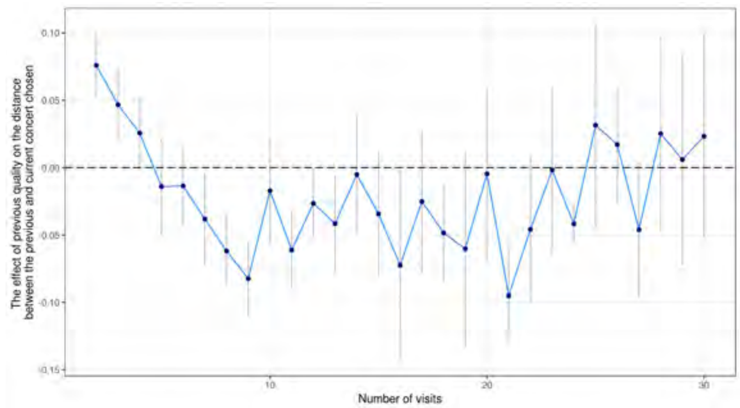
Figure 18 shows the results of the same regression (Equation 5) using the second specification (Gower distance). The results from the second specification give a very similar pattern with the results from the first specification (reported in Figure 8).

A.6 Price variation

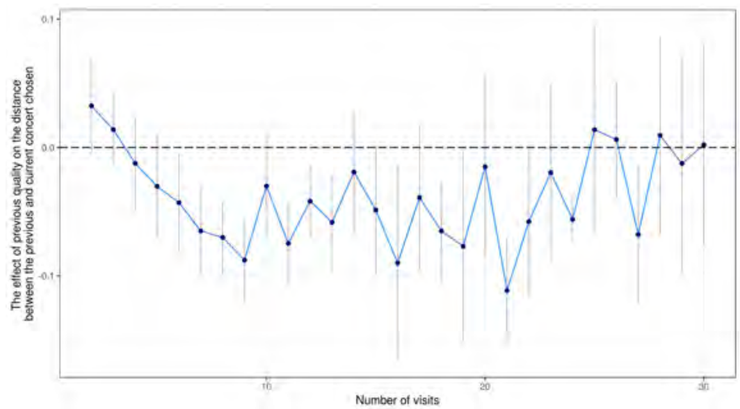
Figure 19 shows the correlation between the listed price and the estimated concert values within genre.

A.7 Rationalizing the descriptive findings with traditional Bayesian learning framework

First, the low return rate of customers after a single visit is hard to be explained by Bayesian learning models with forward-looking behavior. In the context that this paper focuses on, brand (the symphony center) offers a *distribution* of diverse experiences (concerts). Any rational forward-looking consumer would need at least two samples to resolve the uncertainty about the distribution. Therefore, 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.



(a) Using the entire sample



(b) Using consumers with 15+ visits

Figure 18: The effect of current quality on the next choice: the distance between the current and the next concert in the product space
(Using Gower distance measure)

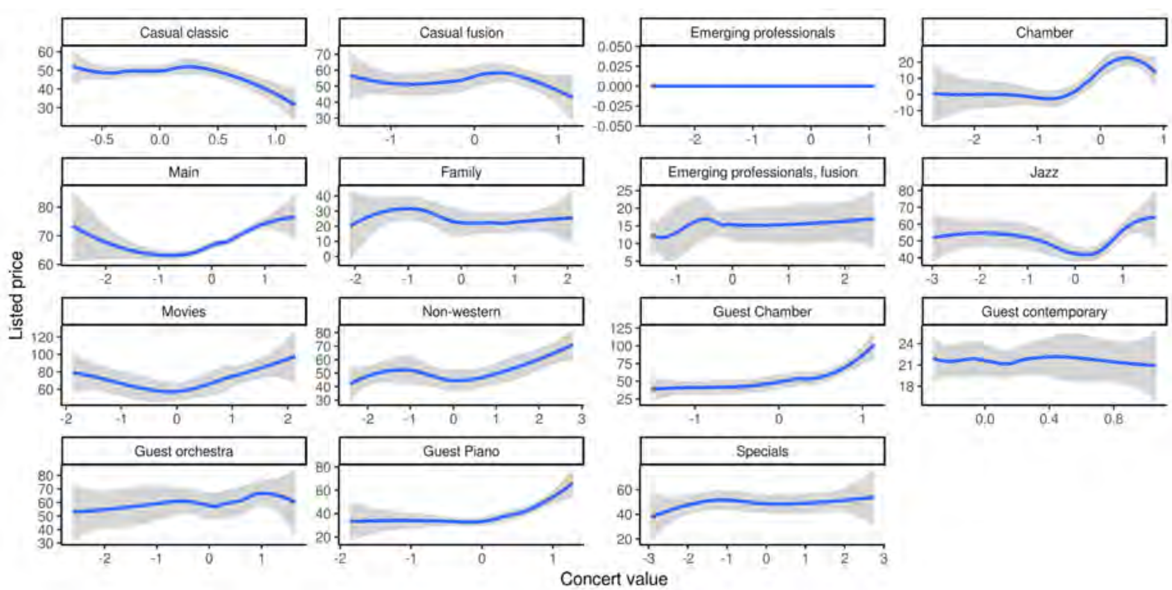


Figure 19: Estimated concert values vs. listed price

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

Suppose we use a Bayesian consumer learning model at the brand level with normally distributed brand 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, motivated by the evidence that there exists a single quality measure for each concert that is correlated with all consumers' choices and return decisions across their levels of experience. 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 and a random error:

$$U_{i1t} = \mu_{it} + \epsilon_{i1t}$$
$$U_{i0t} = u_{i0} + \epsilon_{i0t}$$

where ϵ_{i1t} follows i.i.d. Type 1 Extreme Value distribution and u_{i0} is normalized to be 0. if i decides to consume a concert and realizes a quality draw μ_{it}^s, then she updates her belief 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 the 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 the 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} *(learning)*, or 2) consumer prior already has low mean and variance *(preference heterogeneity in prior)*. 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 very 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, both the *weight* on the signal draws (which is the ratio between prior and signal variance) and the absolute *size* of the signal variance have to be high to fully explain this phenomenon. The size of signal variance, however, has not received much attention to my knowledge, since many empirical works on learning from consumption normalize one of the two variances (signal or prior variance). Figure 20 shows an illustrative example; when the ratio of prior to signal variance is to fixed (2:1) and when both prior mean and signal mean are set to be 0, it is the case of high signal variance which can replicate the high churn rate after a few consumption occasions.

Although able to replicate 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 too restrictive given the empirical context in which customers select the signal by making product choices.

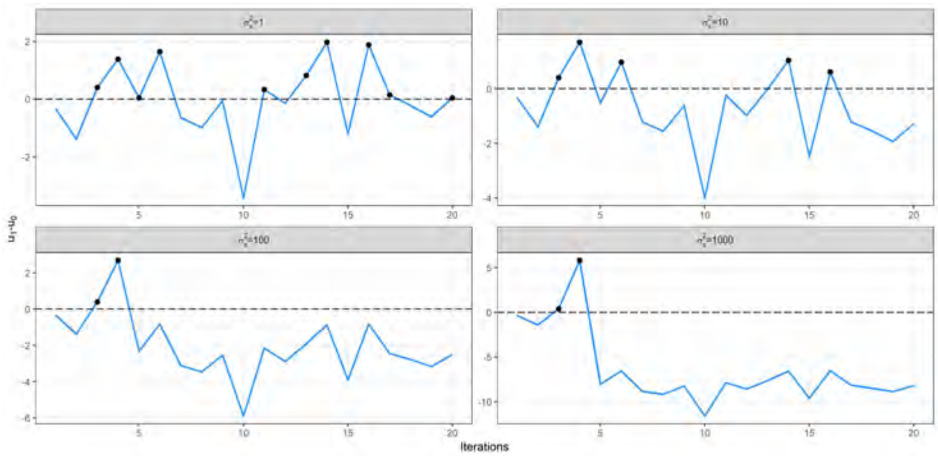


Figure 20: Simulation results: $U_{i1t} - U_{i0t}$. In each simulation, μ_0 and μ_s are both set to be 0, and the ratio of prior variance to signal variance is set to be 2:1. Seed is fixed so that the random errors generated for each period (ϵ_{i0t} and ϵ_{i1t} for all t) are the same across different simulations. 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 brand abandonment.

2) The realized quality in the current visit affects the next product *feature* choice if consumers return. This data pattern can be explained by the learning models with correlated learning across product attributes (Coscelli & Shum (2004), Che et al. (2015)). However, with 120 unique products and 600 different product attributes, modeling product-level or attribute-level correlated learning with separate signal distribution for each of the products or attributes is computationally intractable.

3) The realized quality in the current visit affects the next product *quality* choice if consumers return. As in the previous case, the 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 products, it is challenging for the existing correlated model to explain that consumers choose high quality products even if those new products do not necessarily share the product features with previously consumed products.

1 Estimation results

Table 16: Distribution of posterior means of household coefficients (θ_i): No hetereogeneity

	Min	1Q	Median	Mean	3Q	Max
Concert feature: Time						
Weekend	-4.35	-0.48	0.31	0.31	1.10	5.08
Summer	-9.27	-2.03	-0.95	-0.95	0.13	6.15
Evening	-5.84	-0.25	0.73	0.71	1.68	7.38
Concert feature: Genre						
Others	-13.87	-7.56	-6.31	-6.29	-5.03	1.87
Casual	-12.85	-3.55	-1.96	-1.94	-0.35	8.04
Chamber	-9.04	-2.45	-1.10	-1.11	0.23	7.47
Orchestra	-10.03	-1.81	-0.38	-0.39	1.03	7.60
Family	-13.63	-3.55	-1.56	-1.55	0.43	12.42
Jazz	-8.81	-1.65	-0.23	-0.23	1.19	11.77
Emerging professionals	-14.60	-4.76	-2.91	-2.91	-1.06	9.57
Specials	-6.96	-0.82	0.36	0.36	1.55	8.22
Other utility parameters						
log(Price)	-9.60	-0.11	-0.05	-0.10	-0.03	-0.00
Predicted concert value	0.01	0.43	0.87	1.51	1.79	56.37
Experience spillover						
ρ_0	-5.42	-0.54	0.33	0.33	1.19	5.57
ρ_1	-4.08	-0.79	-0.21	-0.21	0.37	3.50
Search intensity						
ϕ_0	-3.84	-0.14	0.55	0.55	1.23	5.19
ϕ_1	-2.99	-0.08	0.52	0.53	1.12	4.26
ϕ_2	-5.83	-0.42	0.52	0.52	1.46	6.38
ϕ_3	-11.17	-5.60	-4.65	-4.65	-3.70	1.51
Prior knowledge						
ϕ_4	-9.49	-5.76	-5.15	-5.15	-4.54	-1.10

Table 17: Estimation results - Variance (Σ_θ)

	Weekend	Summer	Evening	Others	Casual	Chamber	Orchestra	Family	Jazz	Emerging	Specials	ρ_0	ρ_1	Price	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	δ
Weekend	1.40	-0.28	0.10	-0.65	-0.24	-0.26	-0.26	-0.27	0.17	0.22	0.14	-0.11	-0.32	-0.10	0.09	-0.05	0.08	0.39	0.04	0.22
Summer	-0.28	2.55	-0.28	-0.41	1.89	0.57	1.20	1.20	1.45	1.83	0.15	0.20	-0.06	0.79	-0.18	0.10	1.51	1.14	0.07	0.22
Evening	0.10	-0.28	2.05	-1.58	-0.14	0.36	0.14	0.72	0.17	-0.15	0.23	-0.67	-0.34	0.35	-0.40	-0.23	0.24	0.09	-0.04	-0.25
Others	-0.65	-0.41	-1.58	3.46	-1.59	-2.00	-2.60	-2.07	-2.04	-0.85	-2.25	1.21	0.34	-0.30	0.84	0.72	-0.62	-1.05	-0.11	0.27
Casual	-0.24	1.89	-0.14	-1.59	5.61	1.62	2.63	0.66	0.90	0.86	1.56	-0.96	-0.08	0.04	-0.87	-0.18	0.81	0.42	-0.06	-0.53
Chamber	-0.26	0.57	0.36	-2.00	1.62	3.94	2.64	-0.19	1.30	-0.59	1.60	-0.60	0.47	-0.46	-0.17	-0.47	0.16	0.11	-0.49	-0.04
Orchestra	-0.26	1.20	0.14	-2.60	2.63	2.64	4.41	2.36	1.79	2.21	2.15	-0.54	-0.02	0.60	-0.89	-0.53	0.92	1.22	0.12	-0.03
Family	-0.27	1.20	0.72	-2.07	0.66	-0.19	2.36	8.71	2.05	2.76	1.53	-1.02	-0.03	1.61	-0.87	-1.09	1.58	2.00	0.50	-0.17
Jazz	0.17	1.45	0.17	-2.04	0.90	1.30	1.79	2.05	4.40	1.51	2.10	-0.60	-0.05	0.56	-0.48	-0.60	0.82	1.75	0.66	0.13
Emerging	0.22	1.83	-0.15	-0.85	0.86	-0.59	2.21	2.76	1.51	7.89	0.56	1.30	-0.13	2.50	-1.15	0.59	1.39	2.39	1.21	0.48
Specials	0.14	0.15	0.23	-2.25	1.56	1.60	2.15	1.53	2.10	0.56	3.04	-0.98	0.01	-0.13	-0.79	-0.66	-0.24	0.63	0.31	-0.58
ρ_0	-0.11	0.20	-0.67	1.21	-0.96	-0.60	-0.54	-1.02	-0.60	1.30	-0.98	1.65	0.16	0.33	0.22	0.52	-0.02	0.13	0.19	0.62
ρ_0	-0.32	-0.06	-0.34	0.34	-0.08	0.47	-0.02	-0.03	-0.05	-0.13	0.01	0.16	0.74	-0.15	0.02	0.10	-0.20	-0.22	-0.10	0.02
Price	-0.10	0.79	0.35	-0.30	0.04	-0.46	0.60	1.61	0.56	2.50	-0.13	0.33	-0.15	1.23	-0.37	0.12	0.77	0.92	0.41	0.21
ϕ_0	0.09	-0.18	-0.40	0.84	-0.87	-0.17	-0.89	-0.87	-0.48	-1.15	-0.79	0.22	0.02	-0.37	1.01	0.09	-0.14	-0.37	-0.29	0.27
ϕ_1	-0.05	0.10	-0.23	0.72	-0.18	-0.47	-0.53	-1.09	-0.60	0.59	-0.66	0.52	0.10	0.12	0.09	0.78	0.03	-0.18	0.03	0.13
ϕ_2	0.08	1.51	0.24	-0.62	0.81	0.16	0.92	1.58	0.82	1.39	-0.24	-0.02	-0.20	0.77	-0.14	0.03	1.96	1.12	-0.05	0.41
ϕ_3	0.39	1.14	0.09	-1.05	0.42	0.11	1.22	2.00	1.75	2.39	0.63	0.13	-0.22	0.92	-0.37	-0.18	1.12	2.00	0.45	0.41
ϕ_4	0.04	0.07	-0.04	-0.11	-0.06	-0.49	0.12	0.50	0.66	1.21	0.31	0.19	-0.10	0.41	-0.29	0.03	-0.05	0.45	0.83	0.10
δ	0.22	0.22	-0.25	0.27	-0.53	-0.04	-0.03	-0.17	0.13	0.48	-0.58	0.62	0.02	0.21	0.27	0.13	0.41	0.41	0.10	1.12