Customer retention under imperfect information

Yewon Kim*

July 2021

Abstract

Many multi-product firms see new customers churning quickly after limited product experiences. The paper examines whether early churn is solely driven by customers' low preferences for a given firm or is affected by incomplete information about available products, using individual-level ticket purchases of classical music concerts at a major U.S. symphony center. The data exhibit patterns consistent with consumer learning, which not only suggest incomplete information about products (concerts) among first-time customers but also give a rational for estimating true consumption utilities of concerts inverting experienced customers' choices. Descriptive analyses show a significant impact of experienced utility at the initial visit on subsequent churn, implying that the initial visit affects a customer's expectations about all future concerts. To explore marketing strategies to reduce such information-driven customer attrition, the paper runs counterfactual analyses on policies that offer targeted marketing to second-time customers after their initial visit. The results suggest that it is challenging to earn back customers with targeted offers after their low initial experiences, emphasizing the importance of introductory marketing and choice architecture in customer relationship management.

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

*Stanford Graduate School of Business, yewonkim@stanford.edu. I am grateful to my dissertation committee, Sanjog Misra, Bradley Shapiro, Jean-Pierre Dubé, and Sarah Moshary, for their unlimited support. I thank Pradeep Chintagunta, Anita Rao, Günter Hitsch, Oleg Urminsky, Øystein Daljord, and all other marketing workshop participants at the Chicago Booth Marketing student seminars, The 2020 AMA-Sheth Foundation Doctoral Consortium, University of Texas at Dallas, University of Minnesota, Twin Cities, Columbia University, NYU, and Stanford University for their helpful feedback. I thank the data provider for all the resources. I am responsible for all remaining errors.

1 Introduction

Many firms face a high number of churn events among new customers. High attrition rate at the early consumption stage, followed by decreasing attrition rate over time, is commonly observed both in academic literature (Fader & Hardie (2007), Fader & Hardie (2010)) and in industry reports (Tan (2016), Trivedi (2017), Gessner (2018), Gessner (2019), Cuffe et al. (2018)). Reports consistently deliver three key facts. First, many consumers do not return after few, if not a single, product experiences although firms offer a large collection of diverse products. Second, this pattern holds in various industries including food delivery services (UberEats, Grubhub), lodging services (Airbnb), over-the-top media service (Netflix), mobile games and other subscription-based services (Headspace, Dollar Shave Club). Third, early churn rates vary widely across firms within the same industry. For example, in the meal-kit market, HelloFresh observes 51% of their customers not returning after the first month, while Blue Apron has 38% of their customers churn after the first month.

Understanding the source of early churn is important for a firm for several reasons. First, high acquisition costs makes quick churn costly and calls for more understanding of how to prevent it (Bauer (2017)). Second, effective marketing strategies to win back customers from early churn depend on what drives it. If most of such churn is driven by consumers' low preferences for the firm's overall offering (i.e., latent type of customers as suggested by Fader & Hardie (2010)), setting lower price or changing product portfolio would effectively decrease the churn rate. However, if a significant portion of quick churn is due to imperfect information about the firm's available products, then retention would be increased via informative advertising or temporary price discounts that nudge marginal customers to explore more products. Third, variance in early churn rates across firms within the same industry suggests that quick churn is not a market-fixed characteristics and has room for improvement.

This paper studies what drives churn among new customers. It examines whether quick churn is solely driven by customers' low preferences for a firm or is also affected by incomplete information about available products, using individual-level ticket purchases of classical music concerts at a major U.S. symphony center. The data contains detailed information on individual customers' concert visits over 13 years, including exact program information, net price paid, seat locations, subscription status, and whether tickets are purchased under any promotions.

The research question is decomposed into two steps: 1) whether incomplete information exists in the market, and 2) whether it drives churn. The first step - to show from the data the existence of consumer learning (incomplete information) separate from fixed heterogeneous preferences (full information) - is challenging (Shin et al. (2012)), partly because traditional approach to identify consumer learning relies on specific model structures that *assume* incomplete information (Ching et al. (2013)). Despite its rich academic and managerial implications, this approach is not suitable to address the main question of this paper as it first has to *show whether* consumers have incomplete information. Therefore, instead of assuming consumer learning, I first run a reduced-form test of whether consumers have fully informed, fixed preferences for available products (concerts).

The test uses opposing predictions of standard models with and without complete information. On one hand, a standard learning model predicts that experienced consumers' choices reflect true consumption utilities that inexperienced consumers get to learn about, and therefore inexperienced consumers' choices become more similar over time to the choices of already-experienced consumers. On the other hand, a standard model of fixed heterogeneous preferences predicts no such over-time evolution of inexperienced consumers' choices to a steady state that resembles already experienced consumers' choices. The data show that, over purchase occasions, concert choices made by a fixed group of inexperienced customers become more similar to the choices made by already-experienced customers, rejecting what the model of fully informed fixed preferences predicts. Although not accepting the learning model (as an alternative hypothesis can never be accepted statistically), the test result confirms that the data cannot be fully explained by fully informed heterogeneous preferences only, and that observed consumer behavior is consistent with what the learning model predicts.

The test result consistent with consumer learning not only gives suggestive data-driven evidence of new customers' incomplete information but also gives grounds for identifying true consumption utilities of concerts using experienced visitors' choices. With the rationale in hand, I estimate concert values (average market preference for each concert) by inverting market shares of concerts among experienced visitors (Berry (1994)). The estimated measure of concert values can reflect vertical quality, horizontal match value or both, as a larger number of purchases by experienced customers can represent higher vertical quality or higher average match value. The measure passes several validity checks, including its positive correlation with Billboard classical music album rankings.

Descriptive analyses show the causal impact of imperfect information on customer attrition. An estimated concert value experienced at the initial visit has a significant impact on the probability of subsequent churn at the symphony center level, even when a rich set of confounding factors are controlled for to rule out alternative explanations. A back-of-the-envelope calculation using the regression coefficients suggests that the churn rate after the initial visit would decrease from 60% from 50% if all first-time visitors attend highest-value concerts, which is 17% decrease in churn. The result, combined with the evidence of incomplete information among the first-time customers, indicates a strong causal effect of a single concert experience on a customer's expectations about all other concerts. This finding is also consistent with consumer psychology theories that suggest a significant impact of initial experiences in various contexts (Tversky & Kahneman (1974), Kardes (1986)).

To explore how a firm can reduce visitor attrition using price promotions or product recommendations, I propose a structural model of consumer learning that incorporates wide learning spillovers in the space of rich product characteristics. The model allows for flexible patterns of learning spillover in high-dimensional space in a computationally tractable way. In the model, customers extrapolate their past experiences to predict the value of other untried products (here, concerts) by taking a weighted average of the past experiences. The model reflects several findings from the data that are not fully captured by traditional consumer learning models. For example, visitors in the data set become more likely to select high-value concerts as they get more experienced even if those concerts consist of new features they have not experienced before, which is not fully rationalized by learning only from consumption. To explain such patterns, the model allows consumers to acquire information on available products through an additional channel besides consumption, which is called "search". Here, search refers to any information acquisition behavior other than consumption, ranging from paying attention to the content of promotional materials to searching for product reviews. Consumer incentives to search is a function of how pleasant previous product experiences are, which creates an incremental impact of prior product experiences on subsequent purchase behavior. Specifically, a satisfying prior experience not only raises the likelihood of returning to the firm but also increases the probability of purchasing high-value products in the next period due to increased search, which generates room for endogenously selected signals about the firm.

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

Counterfactual analyses highlight the importance of introductory marketing that steers new customers towards better first-time experience. Simulation using structural parameters shows that even 70% discount offered on the second visit is not sufficient to match the effect of high-value initial experience on customer retention. Counterfactuals also delineate the potential trade-offs in increasing product variety given the information problem and strong learning spillovers. On one hand, more product variety can increase the symphony center's profit by raising the arrival rate of visitors and satisfying the tastes of broader audience. On the other hand, more variety can decrease the profit by increasing the probability of mismatches between first-time customers and available products, inducing subsequent churn. Simulation shows that removing low-average-match-value concerts can raise both ticket revenue and the average number of visits, underlining the negative impact of a large choice set on customer retention due to information problem. Overall, counterfactual exercises imply that firms, when setting marketing strategies, should explicitly consider that customers may rely on the very first experience to determine their subsequent relationship with the firms given the limited information about available products. They also suggest that firms should pay more attention to informing their new customers instead of focusing only on their loyal return customers (Rust et al. (1999)).

The paper complements the existing literature on drivers of customer churn by applying consumer learning models. Customer churn or retention 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 customers churn (Ascarza (2018)). Moreover, to my knowledge, there is no empirical research that looks at why such a high number of churn events take place at the early consumption stage, although most statistical models that predict churn fully take this pattern into account when fitting the data (Fader & Hardie (2007), Fader & Hardie (2010)). Using micro-founded consumer utility model, I view churn as an explicit outcome of consumer learning and search behavior, which offers useful insights on how to design marketing interventions to prevent churn (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 literatures in various disciplines discuss potential welfare loss triggered by incomplete information (Nelson (1970), Stiglitz (1989), Denrell & March (2001), Israel (2005)).

The paper complements the literature on consumer learning. It shows reduced-form evidence of learning by testing different predictions under the model with and without consumer learning. It proposes a new framework that allows for flexible learning spillovers as well as an additional endogenous information acquisition activity without . It also illustrates how the initial consumption experience can have a large enough weight on people's beliefs to determine the entire subsequent consumption 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 discusses the gap between what consumers learn about and which signals they use to learn about it. Facing rapidly changing choice sets, consumers extract information about a firm's entire product offerings by sampling one or few of its diverse products, rather than learning about a relatively homogeneous group of products by trying one of them. This implies that the signals consumers receive and the construct they learn about no longer align perfectly, and that experience spillovers via correlated learning may take place very strongly and broadly (Erdem (1998), Coscelli & Shum (2004), Sridhar et al. (2012), Szymanowski & Gijsbrechts (2012), Che et al. (2015), Ching & Lim (2019)).

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

2 General setting - A market with imperfect information

The paper focuses on the market with two sources of incomplete information. First, consumption utility from a product (vertical qualities, match values, or both) is not fully observable at the purchase stage. Examples include any markets with experience goods (Nelson (1970)). Second, given the large number of products offered by a single firm, consumers do not have full information on the range of consumption utilities that a firm offers with different products. Examples include markets for clothing items, furniture, stationary products, and food items (e.g,. cereals), in all of which each firm provides more than a handful of choice alternatives with numerous varying features.

Consumers engage in two different activities to reduce different types of information gap. To learn consumption utlity of a given product, consumers make purchases and realize the true product value via consumption. Facing a large set of alternatives offered by a single firm, consumers engage in information-seeking activity before purchases to learn which utilities are available, which I call "search." Here, search is not limited to the online context but refers to generic information acquisition behavior besides consumption, ranging from paying attention to product catalogs to reading online reviews.

The setting deviates from the assumptions made in the canonical learning framework in several important aspects. First, the setting differs from the literature in what works as a signal about a firm and how random such signals are. The standard learning model (e.g., Erdem & Keane (1996), Coscelli & Shum (2004), Narayanan & Manchanda (2009), Chintagunta et al. (2009)) assumes that each purchase of any product within a brand gives noisy information ('signal') about a brand (e.g., Tide), which is *randomly* drawn from a distribution with fixed variance relative to consumers' prior. The source of randomness in signals is inherent experience variability that consumers do not have controls for. In this paper, however, each distinct product (e.g., Adidas Ultraboost 5.0 DNA shoes) constitutes a utility distribution offered by a firm (e.g., Adidas) and serves as a signal about the firm, which can be *deterministically* chosen by consumers via product choices. Therefore, the source of the variance in signals is not an experience variability given a product but varying utilities across different products. Second, instead of having a known, fixed variance of a signal distribution in mind, consumers in this setting may have imprecise beliefs about the variance of a signal distribution (the variance in consumption utilities across products), whose beliefs can change via more exposures to different products (Zhao et al. (2011), Dew et al. (2020)). Finally, consumers in this setting can seek out more information on product-specific consumption utility before purchase if they have high perceived variance of consumption utilities across different products.

I illustrate with a simple model how the information frictions can result in quick churn of customers in their early consumption stage. According to the model, consumers churn quickly after a single product experience if they underestimate the variance of consumption utilities from different products, which leads them to 1) choose a product without much information-seeking in their initial purchase, and 2) update their beliefs with high weight on the very first (negative) experience.

2.1 A simple model of learning and search

Suppose there are $J+1$ available products in the market: $j \in \{0, 1, \ldots, N\}$ where $\mathcal{J} = \{1, 2, \ldots, N\}$ is offered by a single firm J and 0 denotes an outside option. After purchasing product j in period t, a risk-averse consumer i has the following consumption utility u_{ijt}:

$$u_{ijt} = v_j - \gamma v_j^2 + \epsilon_{ijt} \quad \forall j \in \mathcal{J} \tag{1}$$

$$u_{i0t} = \epsilon_{i0t} \tag{2}$$

where v_j is the mean consumption utility of product j, $\gamma \geq 0$ is the risk coefficient, and $\epsilon_{ijt} \sim N(0, \sigma_\epsilon^2)$ is an idiosyncratic utility shock that is fully observable to consumers before the purchase.

True mean consumption utility v_j is not observed before consumption, so consumers make purchase decisions based on expected consumption utilities. Expected utility from j is a function of the information set that consumer i has at time t: $I_{it} = \{B_{it}, s_{it}\}$. B_{it} denotes i's belief about the values of available products j at time t that has been formed with previous consumption experiences. For simplicity, I assume that search is a binary activity that reveals true mean consumption utility of a given product if conducted. s_{ijt} is 1 if additional information on product j is obtained via search, and 0 otherwise. Before making a purchase decision, consumer i first decides whether or not to search to acquire additional information on available products based on her belief from prior experiences (B_{it}) and search cost (c_{it}).

B_{it} consists of three types of information on the distribution of consumption utilities offered by firm J: 1) expected mean ($\widetilde{v}_{it}$) of the utility distribution offered by firm J, 2) expected variance ($\widetilde{\sigma}_{it}^2$) of the same distribution, and 3) exact consumption utilities that have been realized via past purchases (v_j). This information set is updated as a function of i's prior belief and consumption experiences. Let $\widetilde{v}_{it}$ and $\widetilde{\sigma}^2$ denote the expected mean and variance of the utility distribution offered by firm J which jointly follow the normal inverse-gamma distribution (Zhao et al. (2011)), and let v_{ijt} denote the perceived value of product j by i at time t. Then, the information set is updated with consumption via the following rules:

$$v_{ijt} = E[v_{ijt}|B_{it}] + \zeta_{ijt} \sim N(E[v_{ijt}|B_{it}], Var[v_{ijt}|B_{it}]) \tag{3}$$

$$E[v_{ijt}|B_{it}]|\widetilde{\sigma}^2_{it} \sim N(\widetilde{v}_{it}, \widetilde{\sigma}^2_{it}/\tau_{it}), \quad 1/\widetilde{\sigma}^2_{it} \sim \Gamma(\alpha_{it}, \beta_{it}) \tag{4}$$

$$\widetilde{v}_{it} = \widetilde{v}_{it-1} + \sum_{j=1}^{3}\left[\frac{D_{ijt}}{\tau_{it-1}+1}(v_j - \widetilde{v}_{it-1})\right] \tag{5}$$

$$\tau_{it} = \tau_{it-1} + \sum_{j=1}^{3} D_{ijt} \tag{6}$$

$$\alpha_{it} = \alpha_{it-1} + \frac{\sum_{j=1}^{3} D_{ijt}}{2} \tag{7}$$

$$\beta_{it} = \beta_{it-1} + \frac{\sum_{j=1}^{3} D_{ijt}\tau_{it-1}(v_j - \widetilde{v}_{it-1})^2}{2(\tau_{it-1}+1)} \tag{8}$$

$$\widetilde{\sigma}^2_{it} = \frac{\beta_{it}}{\alpha_{it}-1} \tag{9}$$

where D_{ijt} is an indicator of purchase of product j; D_{ijt} is equal to 1 if i purchases j at time t and 0 otherwise.

Note that (4) to (8) assume for simplicity that consumers update their beliefs on perceived mean and variance of the utility function only after consumption experiences and not after search. However, the model can be extended to have an additional partial updating after search by adding the modified updating rule in which 1) D_{ijt} is replaced to κs_{ijt} where $\kappa \in [0, 1]$ and 2) $\tau_{it-1} + 1$ is replaced to $\tau_{it-1} + \kappa$. $\kappa = 0$ reduces the model to the baseline model where there is no updating through search; $\kappa = 1$ assumes that information obtained through search affects consumer beliefs in the same way that a consumption experience does.

I assume that the expected mean and variance of consumption utility from product j is the expected mean and variance of the firm J's utility distribution unless i makes search or has purchased the exact same product in the past:

$$E[v_{ijt}|B_{it}] = \widetilde{v}_{it} \tag{10}$$

$$Var(v_{ijt}|B_{it}) = Var(E[v_{ijt}|B_{it}] + \nu_{ijt}|B_{it}) = \frac{\beta_{it}}{\tau_{it}(\alpha_{it}-1)} + \frac{\beta_{it}}{\alpha_{it}-1} = \frac{\tau_{it}+1}{\tau_{it}}\frac{\beta_{it}}{\alpha_{it}-1} \tag{11}$$

$$E[v_{ijt}^2|B_{it}] = Var(v_{ijt}|B_{it}) + E[v_{ijt}|B_{it}]^2 = \frac{\tau_{it}+1}{\tau_{it}}\frac{\beta_{it}}{\alpha_{it}-1} + \widetilde{v}_{it}^2 \tag{12}$$

One can make this assumption more flexible by assuming that 1) consumers form expected utility of product j given observable product characteristics X, and 2) updating of beliefs occurs on the perceived mean and variance of the utility distribution conditional on X.

Then, expected consumption utility of product $j \in \mathcal{J}$ with and without search (before search) can be written as follows:

Expected consumption utility without search:

$$\begin{aligned} E[u_{ijt}|B_{it}, s_{ijt}=0] &= E[v_{ijt}|B_{it}] - \gamma E[v_{ijt}|B_{it}]^2 - \gamma Var(v_{ijt}|B_{it}) + \epsilon_{ijt} \\ &= \widetilde{v}_{it} - \gamma\widetilde{v}_{it}^2 - \gamma\frac{\tau_{it}+1}{\tau_{it}}\widetilde{\sigma}_{it}^2 + \epsilon_{ijt} \end{aligned} \tag{13}$$

Expected consumption utility with search (before search):

$$\begin{aligned} E[u_{ijt}|B_{it}, s_{ijt}=1] &= E[v_j|B_{it}] - \gamma E[v_j|B_{it}]^2 + \epsilon_{ijt} - c_{ijt} \\ &= \widetilde{v}_{it} - \gamma\widetilde{v}_{it}^2 + \epsilon_{ijt} - c_{ijt} \end{aligned} \tag{14}$$

where $c_{it} \sim N(c, \sigma_c^2)$ is search cost. Then, consumer i engages in search if and only if the following two inequalities hold:

$$
\begin{aligned}
& E[u_{ijt}|B_{it}, s_{ijt}=1] > u_{i0t} \\
& \Leftrightarrow \widetilde{v}_{it} - \gamma \widetilde{v}_{it}^2 - c > \epsilon_{i0t} - \epsilon_{ijt} + \eta_{ijt} \qquad (15) \\
& E[u_{ijt}|B_{it}, s_{ijt}=1] > E[u_{ijt}|B_{it}, s_{ijt}=0] \\
& \Leftrightarrow \gamma \frac{\tau_{it}+1}{\tau_{it}} \widetilde{\sigma}_{it}^2 - c > \eta_{ijt} \qquad (16)
\end{aligned}
$$

where $\eta_{it} \sim N(0, \sigma_c^2)$ is the idiosyncratic component of search cost.[1] The first inequality states that the expected utility with search should exceed the consumption utility of an outside option for i to do search; the second inequality implies that the expected consumption utility with search should be greater than that without search.

Let $\phi = Pr(s_{it}=1|B_{it}) = Pr\left(\epsilon_{i0t} - \epsilon_{ijt} + \eta_{ijt} < \widetilde{v}_{it} - \gamma \frac{\tau_{it}+1}{\tau_{it}} \widetilde{v}_{it}^2 - c \text{ and } \eta_{ijt} < \gamma \widetilde{\sigma}_{it}^2 - c\right)$ denote the probability that i engages in search in time t.[2] Given the two inequalities specified above, we can write the following relationship between i's belief (B_{it}) and search probability:

$$
\frac{\partial \phi}{\partial \widetilde{v}_{it}} \propto 1 - 2\gamma \frac{\tau_{it}+1}{\tau_{it}} \widetilde{v}_{it} \begin{cases} \geq 0 & \text{if } \widetilde{v}_{it} \leq \frac{\tau_{it}}{2\gamma(\tau_{it}+1)} \\ < 0 & \text{if } \widetilde{v}_{it} > \frac{\tau_{it}}{2\gamma(\tau_{it}+1)} \end{cases} \qquad (17)
$$

$$
\frac{\partial \phi}{\partial \widetilde{\sigma}_{it}^2} > 0 \qquad (18)
$$

(17) implies that search probability has an inverted V-shape with respect to the perceived mean of the utility distribution, and that the perceived mean is a function of the product utilities that the consumer has experienced so far (based on (4) and (5)). Where the kink occurs in search probability depends on how risk-averse each consumer is; the more risk-averse a consumer is (i.e., higher γ), the lower the threshold of perceived mean utility is where the kink occurs. (18) implies

[1] Given the general definition of search, the source of variation in search cost can include informative marketing campaigns, coverage in media, and other consumer-time-specific shocks that change the level of information exposure.

[2] $\Phi(\cdot)$ denotes standard normal cdf (given the specification in (15) and (16)).

that search probability increases with the perceived variance of the utility distribution ($\widetilde{\sigma}_{it}^2$), and that the perceived variance is a function of the variance of product utilities that the consumer has experienced so far (based on (4) and (8)).

Next, I assume specific values for the model parameters and demonstrate how early customer attrition can occur due to imperfect information.

2.2 An illustration of customer churn under imperfect information

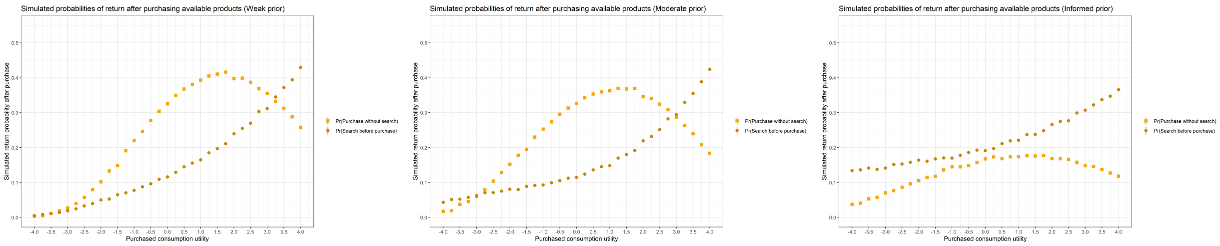
Table 1 summarizes model parameters used for illustration. I simulate product-specific purchase probabilities and return-to-the-firm probabilities in the following period assuming three different consumer priors: 1) weak prior that has low certainty about the mean and underestimates the variance of the utility distribution, 2) moderate prior with the same underestimated variance but with more certainty on the mean of the distribution, and 3) informed prior with more accurate and certain belief on both mean and variance of the utility distribution.

Table 1: Model parameters for illustration

Parameter	Value	Description
Prior and mean search cost $\{\widetilde{v}_{i0}, \alpha_{i0}, \beta_{i0}, \tau_{i0}\}$	{0, 2.1, 0.3, 1} (weak prior) {0.1, 2.1, 0.3, 2} (moderate prior) {0, 4, 9, 5} (informed prior)	Prior belief on mean and variance of the utility distribution
c	0.5	Mean search cost
$\sigma_\epsilon^2, \sigma_c^2$	1	Variance of random shocks in consumption utility and search cost
γ	0.2	Risk parameter
v_j	$\{-4, -3.5, \ldots, 3.5, 4\}$	True mean consumption utility of each product

After computing perceived mean and variance of the utility distribution using (4) to (9), I use the following formula derived from (15) and (16) to simulate probabilities of subsequent purchase and search:[3]

[3]Probabilities in (19) and (20) do not have a closed form solution, so I simulate them with 10000 draws for each random component and report the results.



(a) Weak prior **(b)** Moderate prior **(c)** Informed prior

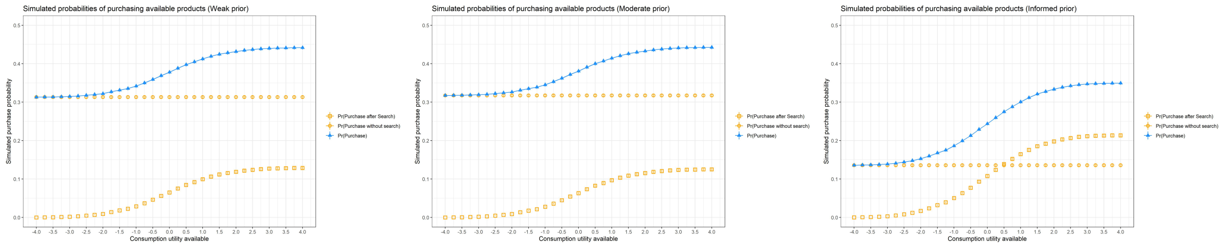
Figure 1: Return probability after experiencing different consumption utilities, by priors

$$\Pr(\text{purchase in } t+1 \text{ without search}) = Pr\left(\epsilon_{i0t} - \epsilon_{ijt} < \widetilde{v}_{it} - \gamma\widetilde{v}_{it}^2 - \gamma\widetilde{\sigma}_{it}^2 \text{ and } \eta_{ijt} > \gamma\widetilde{\sigma}_{it}^2 - c\right) \tag{19}$$

$$\Pr(\text{search in } t+1) = Pr\left(\epsilon_{i0t} - \epsilon_{ijt} + \eta_{ijt} < \widetilde{v}_{it} - \gamma\widetilde{v}_{it}^2 - c \text{ and } \eta_{ijt} < \gamma\widetilde{\sigma}_{it}^2 - c\right) \tag{20}$$

Figure 1(a) shows how an initial product experience can lead a consumer with incomplete information to abandon the firm entirely in two ways. First, the perceived mean decreases dramatically after the initial consumption experience, which makes the probability of return purchase without search close to 0 although there is a high-utility item offered by the same firm. Second, the low perceived mean also reduces the probability of subsequent search to be close to 0, as the search probability is a function of perceived mean as well (Eq. (20)). Even a high perceived variance after the initial low experience, which has a positive effect on search probability (Eq. (18)), cannot raise search probability high enough because of low expected mean utility (Eq. (20)).

Figure 1(b) and (c) demonstrate how the same negative experience can have a much more mitigated effect on churn among experienced customers. Although the probability of purchase without search is low after negative experience ($v_j < -3$) in both cases, the probability of search stays high for customers with more information about the true utility distribution (Figure 1(c)) and is non-zero even for customers with moderate prior (Figure 1(b)). Non-zero search probability implies non-zero probability of finding a high-utility item and purchasing it. This suggests that



(a) Weak prior **(b)** Moderate prior **(c)** Informed prior

Figure 2: Purchase probability of different consumption utilities, by priors

complete churn becomes much less likely when the same low-utility experience takes place later in the consumption journey than earlier, because consumers have more data points from their own consumption experiences to gauge the true mean and variance of the utility distribution and stay willing to search.

Figure 2 demonstrates an additional effect of consumption utilities on churn via deterministic product choices. According to Figure 2(a), consumers with weak prior have high probabilities of buying low-utility products due to lack of information, which reduces subsequent probability of return close to 0. However, consumers with more information are not only less affected by negative experiences (Figure 1(c)) but also less likely to purchase low-utility products given their active search before purchase (Figure 2(c)).

To summarize, when consumers do not have perfect information on both mean and variance of the utility distribution offered by a firm, early consumption experiences affects their decisions to return in two ways. First, the experience directly enters the information set based on which a purchase decision is made. Second, it affects the amount of additional information obtained through search, which also enters the information set for purchase. These direct and indirect effects of product experiences can generate a snowball effect on customer retention. Positive consumption experience not only directly raises the probability of returning to the firm but also can raise the probability of purchasing a high-value item from the same firm due to the incremental information from search. Similarly, negative consumption experience not only directly decreases the probability of returning to the same firm but also decreases the probability of search, which deters consumers from finding the "right" products from the firm even if the highest-value product is indeed offered by the same firm. This rationalizes why and how a single product experience can exert a lasting influence on a customer-firm relationship, even leading to complete churn.

2.3 Predictions of the model

Next, I delineate the model predictions on consumer choices, some of which are empirically tested in the later sections.

H1-1: First-time customers' product choices are not strongly correlated with the underlying consumption utilities. The model predicts randomness of first-time customers' product choices with respect to true consumption utilities for two reasons. First, new customers do not have information on the true consumption utility from each alternative because of lack of consumption experiences. Second, new customers do not have information on the variance of consumption utilities offered by a firm, which affects their search activity before making initial purchases. In particular, underestimation of variance lowers the incentives to search before purchase and makes their product choices less correlated with the true consumption utilities.

H1-2: Experienced customers' product choices are strongly correlated with the underlying consumption utilities. Like any standard learning model, the model predicts that consumers would become fully informed about true consumption utilities once they have enough purchase experiences. More consumption experiences mean more realized consumption utilities which by itself makes consumers more informed. In addition, past purchases also tell consumers about the variance of utilities offered by a firm. If consumers find the experienced variance to be high across their past purchases, consumers become more likely to search and get further product-specific information before making subsequent purchases among untried items. If the experienced variance in utilities from many previous purchases is low, their consumption experiences give highly relevant signals for other untried products. In either way, sufficient consumption experiences make customers well-informed about true consumption utilities. Therefore, experienced consumers' product choices become more correlated with the underlying true consumption utilities.

H2 (*A test for stability of consumer belief*): Inexperienced consumers' product choices become more like experienced consumers' product choices over time. The model predicts over-time evolution of product choices by inexperienced customers, in a way that their product choices become more concentrated around what are chosen by already experienced customers. As inexperienced customers accumulate more consumption experiences, they not only learn about true consumption utilities of the purchased products but also realize the variance of consumption utilities offered by a firm, which makes them more informed about true consumption utilities over time. The model also predicts that customers with enough consumption experiences would be well-informed about the true consumption utilities of products for the same reason (H1). Therefore, the model implies that inexperienced customers' and already-experienced customers' choices become more alike over time as new customers accumulate experiences in the market.

Note that this prediction holds for any standard learning model that assumes convergence of consumer beliefs to true consumption utilities. However, this prediction does not hold under the model of fully informed, time-invariant preferences; under the model of fixed preferences, there should be no systematic evolution of consumer choices when a new consumer cohort's product choices are compared over time to an already experienced cohort's product choices. These opposing predictions by two different classes of model makes it a reduced-form test for whether the data set of interest rejects fixed, fully-informed preferences.

H3: Many first-time consumers do not return at all after a single low-utility experience. As illustrated with the simulation results (Figure 1(a)), the model rationalizes why we observe a significant rate of churn among early-stage consumers after a single or limited product experiences in various markets (Fader & Hardie (2007), Fader & Hardie (2010), Tan (2016), Trivedi (2017), Gessner (2018), Gessner (2019), Cuffe et al. (2018)). Quick churn behavior can be explained by consumers' underestimation of the variance of true utility distribution; this leads consumers to put a much larger weight on a single experience when updating their belief on the mean of the utility distribution than the weight they might have put with correct information on the variance of the signal distribution.

H4: Consumers become better at choosing a high-utility product even among new products they have not experienced before. Search activity outlined in the model allows consumers to become better informed about product-specific consumption utilities at the purchase stage even when faced with completely new alternatives to choose from. As consumers experience variance in consumption utilities across different products provided by a single firm, their likelihood of search before purchase goes up, which leads them to select a high-utility item even among new options they have not experienced before (like in Figure 2(c)). For example, after learning about how varying a product fit can be through their shopping experiences, experienced customers may pay more attention to product reviews when they buy a new clothing item online or even choose to try it on offline before making a purchase. These additional search activities result in better product choices among a new set of alternatives, which cannot be fully explained by learning through consumption only.

After discussing the data used in Section 3, I apply and test these model predictions empirically in Section 4 to study whether part of quick churn can be explained by incomplete information.

3 Empirical setting - A market for classical music concerts

The paper uses individual ticket purchases from a major U.S. symphony center to study the effect of imperfect information on customer attrition. 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 main data set is an individual-level panel of ticket purchases for 13 fiscal years. Detailed information on each concert is extracted from program catalogs, which includes but is not limited to pieces performed, performers, soloists, and solo instruments.

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

3.1 Purchase data and consumer demographics

Table 2 reports descriptive statistics of the ticket purchase data. The data set contains purchases from FY2005 to FY2018, but for the analysis I use purchases from FY2009 to FY2015 to remove potential biases from left truncation.[4] Given 4 years of burn-in and burn-out periods, new customers are identified as those who have not visited the symphony center at least in the past 4 years. Similarly, customer churn is defined as an occasion in which a customer does not return to the symphony center at least in the next 4 years. About 150,000 purchases are made each year by more than 48,000 unique customers. 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.

Price per ticket ranges from \$0 to \$350, and there is a significant amount of within-individual variance in prices paid across different concerts (Table 2(b)). These statistics are based on the

[4]Although the maximum number of days between two purchases by the same customer 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 left truncation bias when labeling customer entrance or departure.

actual transaction prices paid by customers, which vary based on the seat selections and music categories. (See Table 3).

Other information on ticket purchases include exact seat locations, number of tickets ordered, ticket sales channel (e.g., box office, online purchase), price promotion, ticket order date, performance date and time, and whether tickets are purchased as a bundle or as a single ticket. Ticket bundles (called subscription) consist of 3 to 4 individual concert tickets offered at a discounted rate. Note that, although most bundles are configured by the symphony center, the degree of flexibility in bundle consumption is very high given that 1) there are 70 to 120 different ticket bundles that are offered every season and 2) customers can always build their own customized bundles with quantity discount applied.

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

3.2 Concert features

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

There are 14 categories of concerts created by the symphony center. These categories vary in a list of features including the age of pieces performed (e.g., contemporary, classical), genres (e.g., Jazz, movie sound tracks, classical pieces), musical composition (e.g., chamber, orchestra, solo), ambiance (e.g., casual and experimental, traditional and classic), whether the guest performers are invited, whether student artists perform, and specific target audience (e.g., family-friendly). Each category has different baseline price (Table 3).

There are two major sources of variations in concert prices: 1) categories (according to the dispersion in the mean prices in Table 3), and 2) seat locations (according to the average within-

Table 2: Descriptive statistics

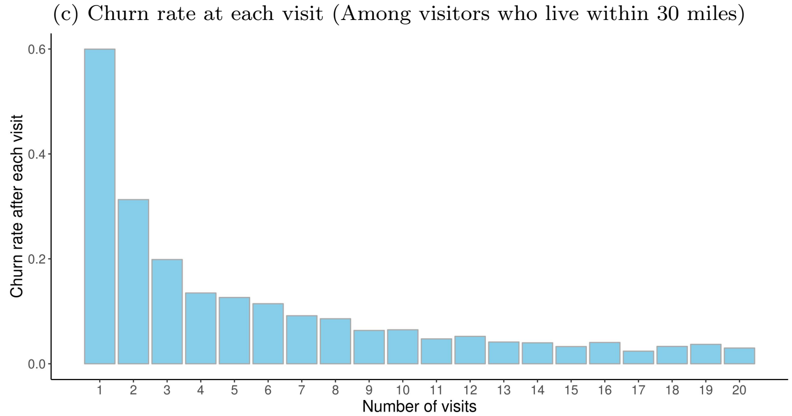
(a) By fiscal year

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 (within 30 miles; 61.3% of the total customers)

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

(c) Churn rate at each visit (Among visitors who live within 30 miles)



concert standard deviation in Table 3). Price also varies across concerts within categories, but not as much as it varies across seat locations or across categories.

Program features used for analysis are summarized in Table 4. 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 3: Category-level information

Category	Unique concerts	Mean price	Standard Deviation in price	
			All (Across-concert + Within-concert)	Average Within-concert
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

Table 4: 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

4 A test for fully informed fixed preferences and identification of true consumption utilities

As the goal of this paper is to study whether incomplete information plays any role in driving early churn, careful examination on whether incomplete information exists in the market should be preceded before investigating its effect on churn. Therefore, instead of assuming consumer learning, I first run a reduced-form test of whether consumers have fully informed, fixed preferences for available products (concerts).

4.1 A reduced-form test for stability of consumer beliefs (H2)

I use the following two assumptions to construct hypotheses that test the time-invariance 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 ν-th visit (purchase occasion) is

$$u_{ij\nu} = \alpha_i + Q_{ij\nu} + \epsilon_{ij\nu}$$

and her choice probability is

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

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

Let $\mathcal{E} = \{i | \nu_i \geq \underline{\nu}\}$ denote a fixed group of experienced customers who have made at least

$\underline{\nu}$ purchases at the beginning of the data period. Let $\mathcal{I} = \{i|\nu_i < \underline{\nu}\}$ denote a fixed group of inexperienced customers who make less than $\underline{\nu}$ by the end of the data period. By construction, the two groups do not overlap.

$\boldsymbol{s}_{\mathcal{E}}$ is a $J \times 1$ vector of observed market shares of products within group $\mathcal{E}$, and $\boldsymbol{s}^{\nu}_{\mathcal{I}}$ represents the same market shares within group $\mathcal{I}$ in their ν-th purchase. This representation assumes that $\mathcal{E}$ has already finished their learning and therefore demonstrates stable product choices across purchase occasions, whereas $\mathcal{I}$ is still in the process of learning and therefore their product choices evolve as a function of the number of purchases.

Using the assumptions and definitions above, I construct the following hypotheses using conflicting predictions of different models:

H_0 *(Stable consumer belief)*: Correlation between $\boldsymbol{s}_{\mathcal{E}}$ and $\boldsymbol{s}^{\nu}_{\mathcal{I}}$ stays the same across I's purchase occasions (ν):

$$\frac{\partial \; corr(\boldsymbol{s}^{\nu}_{\mathcal{I}}, \boldsymbol{s}_{\mathcal{E}})}{\partial \nu} = 0.$$

H_1 *(Unstable consumer belief)*: Correlation between $\boldsymbol{s}_{\mathcal{E}}$ and $\boldsymbol{s}^{\nu}_{\mathcal{I}}$ changes over I's purchase occasions (ν):

$$\frac{\partial \; corr(\boldsymbol{s}^{\nu}_{\mathcal{I}}, \boldsymbol{s}_{\mathcal{E}})}{\partial \nu} \neq 0.$$

To test the hypotheses, I use the following step:

1. I create $\mathcal{E} = \{$a sample of 5000 experienced customers who have made at least 10 past visits at the beginning of the data period$\}$.[5] Similarly, I create $\mathcal{I}_{\overline{\nu}} = \{$a sample of customers who make total $\overline{\nu}$ visits in the data period$\}$. I try different values $\overline{\nu}$ to capture potentially very different sets of new customers: $\overline{\nu} \in \{2$ (new customers who eventually visited only 2 times during the data period), $10+$ (new customers who eventually visited more than 10 times during the data period)$\}$.

2. For each concert, construct $s_{\mathcal{E}j}$ that represents what percentage of experienced visitors choose

[5] 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.

concert j among all concerts in the same year. Specifically, $s_{\mathcal{E}j}$ for concert j is calculated as

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

where $y_{ij} = 1$ if consumer i purchases tickets for concert j. For example, if both concert A and concert B take place in 2008 and $s_{\mathcal{E}A} = 0.1$ and $s_{\mathcal{E}B} = 0.4$, it means that concert B is more favored by experienced visitors than A as 10% of the experienced visitors in 2008 choose A while 40% choose B.

3. Test whether visitors in $\mathcal{I}_{\overline{\nu}}$ become more likely to choose a concert with high market shares among experienced visitors ($\mathcal{E}$) over time. To do so, I run the following regression with individual fixed effects (α_i) using concert choices by $\mathcal{I}_{\overline{\nu}}$:

$$(s_{\mathcal{E}j} \text{ of concert } j \text{ chosen by consumer } i \in \mathcal{I}_{\overline{\nu}} \text{ at visit } \nu_i < \overline{\nu}) = \alpha_i + \beta_\nu \nu_i + \eta_{ij} \quad \forall i \in \mathcal{I} \quad (21)$$

and check whether β_v is statistically different from 0.

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

Although bundle purchases are flexible enough and thus less restrictive given the large number of different bundles offered (up to 120 bundles), it may restrict consumers' choices and introduce additional noise to the market shares computed above. As a validity check, I run the same test with and without bundle purchases and confirm that the results do not change. I present the results that only use non-bundle purchases for both experienced and inexperienced customers.

Test results Figure 3 shows the evolution of correlation between the choices by already experienced customers and by inexperienced customers over time ($corr(\boldsymbol{s}_{\mathcal{E}}, \boldsymbol{s}^{\nu}_{\mathcal{I}_{\overline{\nu}}})$). Each line denotes correlation between the vector of concert choices by already experienced customers ($\boldsymbol{s}_{\mathcal{E}}$) and those by inexperienced customers with $\overline{\nu}$ total visits during the data period ($\boldsymbol{s}_{\mathcal{I}_{\overline{\nu}}}$). For each line, customer

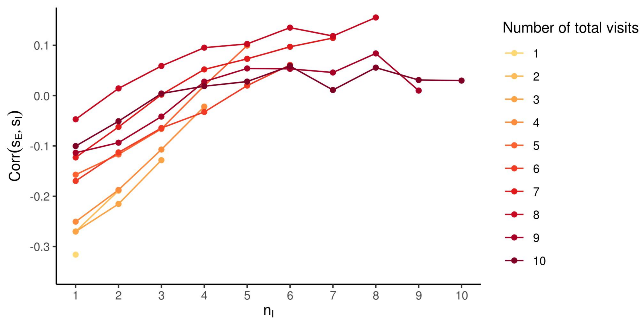


Figure 3: Evolution of correlation between sample experienced ($\mathcal{E}$) and inexperienced consumers ($\mathcal{I}_{\overline{\nu}}$)' choices over time, by the total number of visits of inexperienced consumers during the data period ($\overline{\nu}$)

base stays the same across the number of cumulative visits (x-axis) with no customer entry or exit. Upward-sloping patterns for all fixed groups of customers with different total visits imply that, as inexperienced customers make more visits, their concert choices look more like the choices of already existing experienced customers. Importantly, even the choices made by short-lived visitors - those who do not return to the symphony center after 2 visits - still demonstrate an increasing correlation with already experienced visitors' choices over visits. This implies that the systematic evolution of concert choices is not just a feature of visitors who end up staying in the market for a long time but is found in a broad set of audience.

Table 5 reports the test results. Statistically significant β_ν's in both short-lived and long-lived customer group reject the null hypothesis that inexperienced customers' beliefs stay constant across 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) New consumers who eventually make 2 visits	(2) New consumers who eventually make 10+ visits
β_n (Number of visits)	0.017***	0.011***
	(0.004)	(0.001)
Observations	83,122	45,332
R^2	0.730	0.490
Note:		*p<0.1; **p<0.05; ***p<0.01

Interpretation of the test result The test result provides evidence that consumer beliefs about product-specific consumption utilities are not stable over purchase occasions. There can be various factors that drive such pattern. Consumer learning about product values rationalizes the pattern of changing beliefs. Targeted advertisement that promotes different products to customers based on their length of consumption history can also justify the same pattern. In this specific empirical context, the targeted advertisement effect is ruled out given that most marketing materials provided by the symphony center do not highlight different concerts based on consumer loyalty. Time-varying preferences can also rationalize the observed pattern. However, it is unlikely to expect a new visitor's preference to change to be closer to experienced visitors' after just a single visit, which is observed in the data.

Although the test does not accept consumer learning, it rejects that the data can be explained solely by the model of fixed preferences under perfect information (Remark 1). In other words, the test does not reject the existence of time-invariant heterogeneous preferences in the data; instead, it rules out the non-existence of time-varying components in consumer beliefs. Moreover, the result is consistent with what standard consumer learning models predict (Remark 2). This framework of hypothesis testing can be useful in future empirical research at checking whether the data is suitable for studying consumer learning.

Based on the test results suggestive of consumer learning, I estimate the true consumption utilities of available concerts using experienced consumers' choices.

4.2 Identification of consumption utilities (H1-2)

Under the standard assumption of consumer learning which the test result supports, I estimate the true mean consumption utility of each concert by inverting experienced consumers' choices à la Berry (1994).

Suppose that 1) the random utility component follows i.i.d. Type 1 Extreme Value distribution, 2) consumers make a decision to choose up to one concert in a given week,[6] and 3) there is no random coefficients within the group of experienced consumers ($\mathcal{E}$). Then, $\widehat{Q}_j^*$ measures the underlying concert value Q_j^* (Berry 1994):

$$\begin{aligned}
&\widehat{Q}_j^* = \ln(s_{\mathcal{E}j}) - \ln(s_{\mathcal{E}j0}) - \text{(Hour, Day, Month, Genre Fixed Effects)} \\
&\qquad \text{where} \quad s_{\mathcal{E}j} = \frac{\sum_{i\in\mathcal{E}} y_{ij}}{\sum_{i\in\mathcal{E}} \sum_{k\in\{k|Year_k=Year_j\}} y_{ik}} \quad \text{and} \quad s_{\mathcal{E}j0} = 1 - \sum_{k\in Week_j} s_{\mathcal{E}k} \\
&\qquad \text{and} \quad \mathcal{E} = \{i|i\text{'s number of visits at the beginning of Year}_j = \nu_i \geq \overline{\nu}\}.
\end{aligned} \tag{22}$$

y_{ij} is 1 if consumer i purchases a ticket for concert j, and $Year_j$ indicates the year in which concert j takes place. Here, market size for each concert (that is used to calculate the share of no purchase) is defined by the number of ticket purchases made by experienced customers within the sample group ($\mathcal{E}$) in the same year. As discussed in Section 2 and 3, I include both bundle- and non-bundle purchases given the large number of available bundles (about 100 different bundles per season) and flexible options to create own bundles.

This estimator, computed with an assumption of no heterogeneity, identifies the market's true average consumption utility from each concert. More specifically, it reflects the market preference for each concert in a given year in a given genre, voted by a holdout group of already experienced

[6] In other words, consumers are assumed to face a multinomial decision for every week. This assumption is not unreasonable as many loyal customers to the symphony center visit the symphony center more than once per month.

customers, net of time effects that may affect visitor traffic. The estimator can be easily modified to explicitly account for heterogeneity in perceived consumption utilities, which is demonstrated in the Online Appendix. For the subsequent analyses, I use $\widehat{Q}_j^*$ (following Equation (22)) as a proxy for market average preference for concert j.

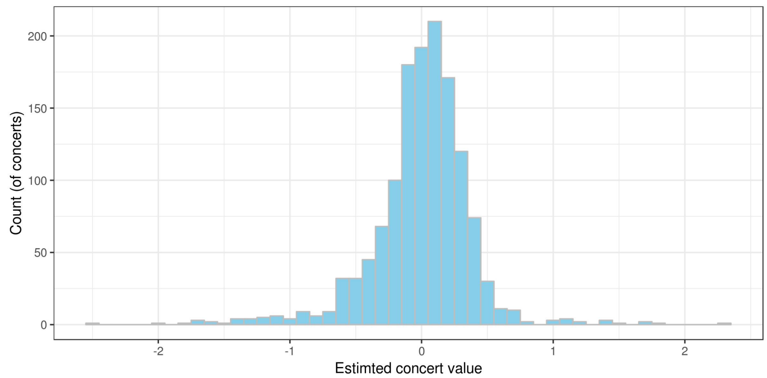


Figure 4: Histogram of estimated concert values

Distribution of the estimated concert values using 5000 experienced consumers with $\overline{\nu} = 15$ is summarized in Figure 4.[7] Section A.1 in the Appendix further discusses the specification of the estimator and its implication. Section A.2 discusses the validity of the estimates, in which I show their positive correlation with billboard rankings of classical music albums as well as with other behavioral responses of visitors that indicate perceived concert qualities.

The idea of using choices by a subset of individuals to recover the true demand values has been applied in prior research across disciplines 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. Anderson et al. (2015) and Simester et al. (2019) use purchases of a certain group of consumers as a predictor of new product failures in various categories. More recently, Huang et al. (2018) and Doraszelski et al. (2018) back out the information that new agents

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

try to learn by using observations on the experienced agents' behavior.

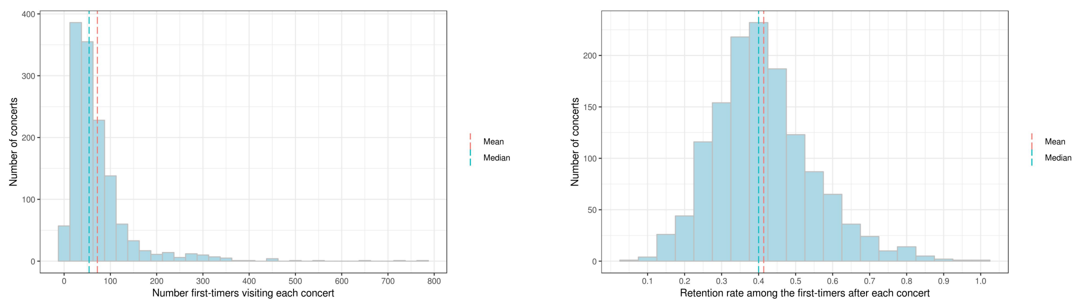
This approach deviates from how most empirical literature demonstrates consumer learning. Instead of jointly estimating consumer preferences and learning parameters based on specific functional form assumptions (e.g., Bayesian updating rule with normal signals and priors), it first estimates consumer preferences under full information using experienced consumers' data and then recover learning behavior by tracing which product values consumers choose over time. Still, identification of true consumption utilities relies on the same standard assumption that consumers' beliefs about consumption utilities converge to the true values as they accumulate more consumption experiences. This approach can be useful when learning is likely to take place but not in a shape that traditional Bayesian learning models predict. The approach can also easily incorporate heterogeneity in perceived product values (See the Online Appendix for more discussion).

5 Descriptive analysis

Using the data set and the estimated true consumption utilities, I show that quick churn is partly due to consumers' incomplete information about available products. In the following descriptive analyses, I include only households within 30 miles from the symphony center based on the zip code information.

5.1 Varying retention rates across concerts

Figure 5 shows a large variation across concerts in both the number of first-time customers (Figure 5(a)) and the rate of subsequent churn among them (Figure 5(b)). Average retention rate after visiting a single concert is about 40%, implying that more than half of the visitors on average do not come back to the symphony center after their first visit. However, the distribution has a long right tail, meaning that there are a set of concerts after which more first-time visitors stay for subsequent visits.



(a) The number of first-timers arriving at each concert

(b) Retention rate among the first-timers after each concert

Figure 5: Variations across concerts in the first-time visitors' arrival and retention rates

5.2 The effect of experienced consumption utilities on customer churn

To analyze the effect of experienced concert values on the following churn, I run a regression with individual-level churn decision as an outcome variable and the recent concert value experienced as the key explanatory variable:[8]

$$Pr(Churn_{i\tau} = 1|\tau\text{-th visit}) = \frac{\exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)}{1 + \exp(\beta_{q1}\widehat{Q}_{i\tau} + \beta_{q2}\widehat{Q}_{i\tau-1} + \beta_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\gamma)} \tag{23}$$

where $\widehat{Q}_{i\tau}$ is the estimated concert value that i experienced at her τ-th visit. Similarly, $\widehat{Q}_{i\tau-1}$ and $\widehat{Q}_{i\tau-2}$ are the lag concert values experienced at $\tau-1$-th and $\tau-2$-th visit respectively. To ease the interpretation, I normalize $\widehat{Q}_{i\tau}$ to have mean 0 and standard deviation of 1. Table 6 describes a set of control variables ($Z_{i\tau}$) included in the regression.

Detailed data in each transaction allows me to control for various alternative explanations that may cause spurious correlation between the experienced concert values and subsequent churn decisions. For example, price paid controls for a potential effect of price on churn decision. The number of days between the date of ticket order and the date of concert absorbs the effect of

[8]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. As I discard the last 4 years of ticket purchase data as a burn-out period, this definition of churn implies that $Churn_i\tau = 1$ if visitor i does not come back to the symphony center after τ-th visit *at least* for the next 4 years. Several other specifications of churn (e.g., churn defined as not coming back for n year with $n \in \{1, 2, 3, 4\}$) give qualitatively the same results.

Table 6: 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 (22) but using the choices of a holdout sample of 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)

ticket availability at the moment of purchase. Concert genres control for potential differences in the underlying churn rates across customers who prefer different genres. A measure of concert popularity among the first-time visitors controls for the baseline difference in the churn rate between one-timers and other customers (i.e., controlling for visitors to those concerts that are only tailored to casual visitors). If there is no causal impact of experienced concert value on churn, then the coefficient of recent concert value (β_{q1} in Equation 23) would become insignificant after controlling for these variables.

Table 7 summarizes the results. Significant and negative coefficients of concert values, even when a list of confounding factors are controlled, imply that there is a systematic relationship between experienced concert values and subsequent churn rates: the higher the experienced concert value is, the lower the chance that a consumer churns after the visit. Exclusion of control variables do not create any qualitative difference in the regression results as Column 4-6 show. The coefficient in Column 1 implies that 1-standard-deviation increase in the first experienced concert value leads to the reduction of average churn rate from 61.3% to 59.5%.

A statistically significant coefficient of the most recent concert value when the lag concert values are included further supports the causal impact of experienced concert value on churn. If we assume that the first concert choice informs a customer's latent type (e.g., long tenure vs. short tenure) and

Table 7: The effect of consumption experience on subsequent churn decision (Binary logit model)

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

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

her churn decision is solely driven by her latent type, controlling for the concert value chosen in the first visit should remove the effect of the most recent concert value on the churn decision. Under this assumption, the persistent effect of the most recent concert value on customer churn even after controlling for the previous concert choices implies the causal impact of experienced concert value on consumer decisions to churn. Different specifications like a linear probability model with even more control variables (e.g., seat locations and ticket sales channel) do not change the results, which are reported in the Online Appendix.

Initial concert experiences also affect other factors that determine customer lifetime value, including the length of tenure or the amount of donation. In Section A.6, I run the same regression as in Equation (23) using two alternative outcome variables: conversion to a regular consumer (with 10+ visits) and conversion to a donor.[9] The results imply that the initial concert experience is correlated with both the probability of being a regular customer beyond the second visit and the probability of making donation to the symphony center.

In summary, the causal impact of experienced consumption utilities on customer retention is supported by two joint pieces of evidence: 1) the incomplete information on concert values among first-time visitors, and 2) the correlation between the experienced concert values and customer churn, which stays significant even when a rich set of vsitor-concert-specific variables are controlled for. A back-of-the-envelope calculation suggests that the average churn rate after the initial concert visit can decrease by 10 percentage point if every visitor is treated with the highest-value concerts on their initial visits. The effect of initial concert experience on customer lifetime value is even greater when its impact on customer tenure and donation is taken into account.

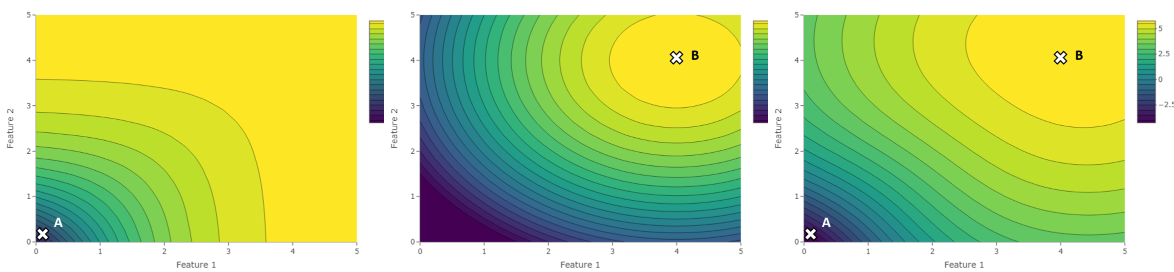
To study how to prevent and reverse churn induced by initial concert experiences via marketing interventions, I propose in the next section a full structural model of visitors' concert choices that extends the core idea of the simplistic model in Section 2.

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

6 A model of consumer learning for a large group of products via multiple channels

I propose a structural model of consumer learning for a large group of products. The model explains flexible patterns of learning spillovers from a single consumption experience to a number of other untried products in a computationally tractable way. It also introduces additional information acquisition channel ("search"), which allows consumers to become better at selecting high-value products among those they have not consumed before.

Consumers extrapolate their past consumption experiences to other untried products by taking the weighted average of those experiences. The weight on each consumption experience is assigned based on the product similarity between the previously experienced product and the untried product of interest.



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

Figure 6: Illustration of learning spillover from consumption

Figure 6 shows an illustrative example of how learning spillovers from consumption experiences take place in a product feature space. Two axes denote two available product features, and different colors denote the range of predicted consumption values for each combination of the two product features. The brighter the color is, the higher the predicted consumption value is. Initially, a consumer tries product A and experiences a low value from it. Given the experience, she updates her prediction on all the other combinations of product features based on their distances from A in the feature space (Figure 6(a)). In the next purchase, she tries product B which is predicted to have a high value based on her previous experience. Suppose that product B indeed gives high-value experience (Figure 6(b)). The resulting belief about all the other untried products after the two purchases is the weighted average of the predicted values in (a) and (b) (Figure 6(c)).

Consumers can obtain information about available products before purchase through an additional channel, which I call search. As discussed in Section 2, decision to search is determined by how pleasant the previous consumption experiences are; the better the prior experiences are, the higher the incentive for search is in the next period. This creates an incremental impact of prior experiences on subsequent purchase behavior, since the amount of product-specific information that customers obtain before making purchase is a function of their past consumption experiences.

6.1 Utility

Each week, consumers choose among concert offerings of the week and an outside option (i.e., not purchasing any concert ticket). Purchase decision maximizes the predicted consumption utility from visiting a concert, which is formed based on the acquired information from the past visits and from search in the current period. I assume that consumers are myopic, i.e., they maximize the current-period predicted consumption utility. [10]

Consumption utility is realized at a concert visit; it is a function of an individual customer's preference over different genres and the true value of the concert visited. Her experience from the visit is then used to update her belief on any upcoming concerts, which is discounted by how different the upcoming concerts are from the one she has visited. The concert experience also determines search intensity in the upcoming weeks by forming her expected consumption and search benefit.

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

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

$$
\begin{aligned}
u_{ijt} &= U(Q^*_j, X_j \epsilon_{ijt}) = \gamma_{ig} + \tau_{ij} + \delta_{i1} Q^*_j + \lambda_i F_j - \alpha_i p_j + \epsilon_{ijt} \\
u_{i0t} &= U(u_0, \epsilon_{i0t}) = u_0 + \epsilon_{i0t}
\end{aligned} \tag{24}
$$

where γ_{ig} is i's preference for genre g that concert j belongs to, τ_{ij} is the month, day, and hour fixed effect, $\widehat{F}_j$ is the popularity of concert j among first-time visitors (constructed in Section 5.3), and p_j is the price of concert j.[11] $\widehat{F}_j$ is included in addition to $\widehat{Q}^*_j$ to capture customers' heterogeneous tastes for specific concerts that are tailored mainly for one-time visitors (e.g., concerts featuring pop music in a botanical garden). δ_i captures i's sensitivity to the estimated true concert value $\widehat{Q}^*_j$. ϵ_{ijt} is the random component of consumption utility that is known to consumer i but is unobservable to researchers. u_{i0t} refers to the utility from consuming an outside option. u_0 is normalized to be 0.

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

$$
\begin{aligned}
\widetilde{u}_{ijt} &= U(\tilde{Q}_{it}(X_j), X_j, \epsilon_{ijt}) = \gamma_{ig} + \tau_{ij} + \delta_{i1} \widetilde{Q}_{it}(X_j) + \lambda_i \widehat{F}_j - \alpha_i p_j + \epsilon_{ijt} \\
\widetilde{u}_{i0t} &= U(u_0, \epsilon_{i0t}) = u_0 + \epsilon_{i0t}
\end{aligned} \tag{25}
$$

where $\widetilde{u}$ denotes the predicted consumption utility and $\widetilde{Q}_{it}(X_j)$ denotes the concert value predicted by consumer i at week t given the observable product characteristics X_j.[12]

Let $\widetilde{v}_{ijt}$ denote the deterministic component of predicted consumption utility in researcher's perspective. Assuming that the random utility component ϵ_{ijt} is *i.i.d.* Type 1 Extreme Value, I can write the purchase probability of concert j among the available concerts at week t ($\mathcal{J}_t =$

[11] To avoid modeling consumers' seat choices, I use concert-level price (p_j) instead of the actual price paid by individual customers (p_{ij}) in this model. Concert-level price for each concert is obtained as the average ticket price for Main Floor seats.

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

$\{0, 1, \ldots, J_t\}$) as

$$Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\widetilde{v}_{ijt})}{\sum_{k \in \mathcal{J}_t} \exp(\widetilde{v}_{ikt})}. \tag{26}$$

6.2 Updating beliefs about concert values through consumption

After visiting concert j, consumer i uses the difference between the experienced concert value and the predicted concert value to update her belief about the upcoming concerts. Let Δ_{ijt} denote the discrepancy between the true and the predicted value based on the previous concert visit experiences :

$$\Delta_{ijt}(Q^*, X_j) = Q_j^* - B_{it}(X_j). \tag{27}$$

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

$$B_{it+1}(X_k) = B_{it}(X_k) + g_{ijkt} \cdot \Delta_{ijt} \tag{28}$$

$$\text{where } g_{ijkt} = (\exp(\rho_{it} \cdot d(X_j, X_k)))^{-1}. \tag{29}$$

Here, similarity between product j and k is known to both consumers and researchers based on the observed product features X_j and X_k.[13] g_{ijkt} determines how much consumer i generalizes the information from a single concert experience (Δ_{ijt}). If $\frac{\partial g}{\partial d} = 0$, then the discrepancy Δ_{ijt} is equally applied to the beliefs of all the untried concerts regardless of how different the concerts are from the concert recently visited (j in this case). If $\frac{\partial g}{\partial d} < 0$, learning spillover from j is stronger for those concerts that are more similar to j. I assume that $|g(\cdot)| \leq 1$, i.e., the weight on the spillover from the recent visit cannot exceed 1. ρ_{it}, the parameter that determines the locality of spillovers in the

[13]The model does not allow heterogeneity across consumers in perceived similarities between products.

concert feature space, is specified as follows:

$$\rho_{it} = \exp\left(\rho_{i0} + \rho_{i1} N_{it}\right). \tag{30}$$

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

6.3 Updating beliefs about concert values through search

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

$$\begin{aligned}
\widetilde{Q}_{it}(X_k) &= B_{it}(X_k) + \phi_{it}(Q_k^* - B_{it}(X_k)) && (31)\\
&= \phi_{it} Q_k^* + (1 - \phi_{it}) B_{it}(X_k) && (32)\\
\text{where } \phi_{it} &= \text{information gain on true concert value via search (``search intensity'')}\\
&= \phi(\text{expected consumption benefit, expected gain from search,}\\
&\qquad \text{the amount of concert information the symphony center sends to } i)\\
&= \phi\left(\frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} Q_j^*,\ \ \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} \left[Q_j^* - \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} Q_j^*\right]^2, N_{it}\right) && (33)\\
\text{and } \phi_{it} &\in [0, 1]
\end{aligned}$$

Search intensity is a function of i's motivation for search, which is endogenously determined by expected consumption benefit and search benefit. It is also affected by the amount of information that the symphony center sends to consumer i via various marketing channels, as search captures all the external sources of concert information besides visits. I assume that the same set of concerts are advertised by the center to all customers, but the advertising intensity or the amount of details provided for each concert can differ across customers.

$\mathcal{C}_{it}$ refers to the set of concerts that i has visited up to week t. Expected consumption benefit is modeled as the average realized concert value from the previous visits up to week t, and the expected gain from search is modeled as the variance of the experienced concert values up to week t. N_{it}, the number of visits, is included in the search intensity parameter because the frequency of targeted advertising to consumer i (e.g., direct mails on upcoming concerts, pamphlets distributed to visitors at the venue) depends on how many visits she has made so far.

I assume that $\phi_{it} \in [0,1]$; $\phi_{it} = 0$ means that there is no further correction of predicted value via search, while $\phi_{it} = 1$ implies that the true value is fully recovered via search. I specify ϕ_{it} as follows:

$$\begin{aligned}
\phi_{it} &= \frac{\exp(\phi_{i0}a_{it} + \phi_{i1}b_{it} + \phi_{i2}N_{it} + \phi_{i3})}{1+\exp(\phi_{i0}a_{it} + \phi_{i1}b_{it} + \phi_{i2}N_{it} + \phi_{i3})} \\
\text{where } a_{it} &= \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} Q_j^* \\
\text{and } b_{it} &= \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} \left(Q_j^* - \frac{1}{N_{it}} \sum_{j \in \mathcal{C}_{it}} Q_j^* \right)^2 .
\end{aligned} \tag{34}$$

Note from Equation (32) 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 concert experiences. Therefore, the search step described here can also be viewed as a learning process with heteroskedastic signals whose variance is endogenously determined by previous consumption experiences.

6.4 Prior knowledge about concert values

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

$$\widetilde{Q}_{i0}(X_j) = \phi_{ip} Q^*_j$$
$$\text{where} \quad \phi_{ip} = \frac{\exp(\phi_{i4})}{1+\exp(\phi_{i4})}. \tag{35}$$

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

6.5 Identification

Which concerts consumers subsequently visit if any is the key to decompose learning from search and learning from consumption when only consumption (ticket purchase) data is available. According to the model, consumption experiences do not allow consumers have better information on other high-value concerts that are not particularly similar to the concerts visited. Learning from search, however, allows consumers to have better information about all untried concerts no matter how similar they are to the ones previously visited. Therefore, learning from consumption is identified if consumers are able to select (avoid) high-value (low-value) products among the concerts similar to the previously consumed, whereas learning from search is identified if consumers are able to select (avoid) high-value (low-value) concerts from an unfamiliar group of concerts.

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

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

7 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.[14] Product distance between any given two concerts is computed via weighted Gower distance (See Section A.7 in the Appendix) and is scaled so that the maximum observed product distance is 1. To estimate

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

the model, I replace Q^* with $\widehat{Q}^*$ which is estimated in Section 4.2.

θ_i denotes a vector of utility parameters that are estimated:

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

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

$$\begin{aligned}
&\theta_i = Z_i\Gamma_{s_i} + \iota_i \\
&\iota_i \sim N(0, \Sigma_{s_i}) \\
&s_i \sim \text{Multinomial}_K(\pi)
\end{aligned} \tag{36}$$

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

Priors are defined as follows:

$$\begin{aligned}
&\Gamma \sim N(\overline{\Gamma}_{s_i}, A^{-1}_{\Gamma_{s_i}}) \\
&\pi \sim \text{Dirichlet}(\overline{\alpha}) \\
&\Sigma_s \sim IW(\nu, V).
\end{aligned} \tag{37}$$

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:

$$\theta_i | s_i, Z_i\Gamma_{s_i}, \Sigma_{s_i} \tag{38}$$

$$\pi, s, \{\Gamma_s\}, \{\Sigma_s\} | \{\theta\}. \tag{39}$$

I use a random-walk Metropolis-Hastings step to draw θ_i given other sets of parameters (Step (38)). To make draws of mixture components (Step (39)), I use `rmultireg` and `rmixGibbs` from bayesm package (Rossi et al. (2005)).

8 Estimation results

Table 8 and Figure 7 report the parameter estimates from the 2 mixture-of-normal model. There is a major group (Group 1; 84.9%) and a minor group (Group 2; 15.1%) with difference in tastes for genres, sensitivity to concert values, and the amount of prior information about the underlying concert values.

Table 8: Distribution of posterior means of household coefficients (θ_i): 2 normal mixture components

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

Figure 8 translates the prior knowledge parameter (ϕ_{i4}) to the fraction of the underlying concert

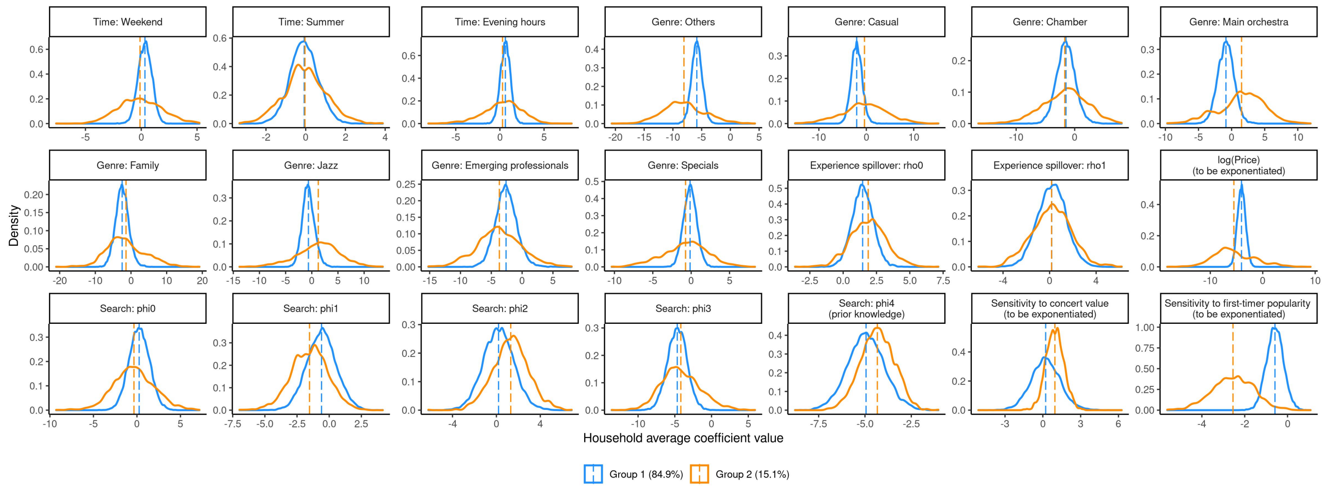


Figure 7: Distribution of posterior means of individual-level coefficients (θ_i): 2 normal mixture components

values known to consumer i before the first visit($\phi_{ip} = \frac{exp(\phi_{i4})}{1+exp(\phi_{i4})}$; Equation 35). Although Group 2 is reported to have more prior knowledge about the underlying concert values with a long right tail, both groups on average have little information about the overall concert values.

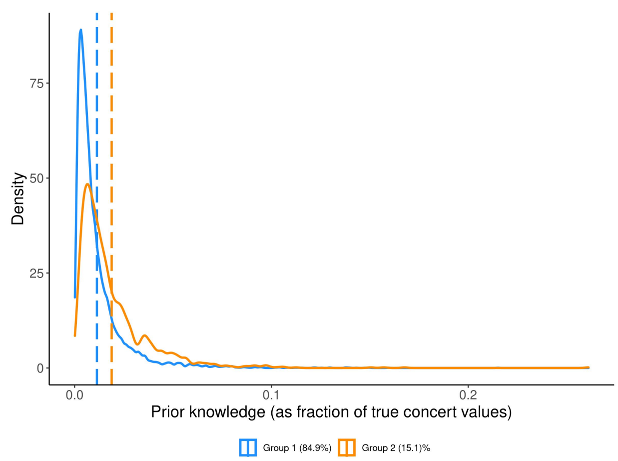
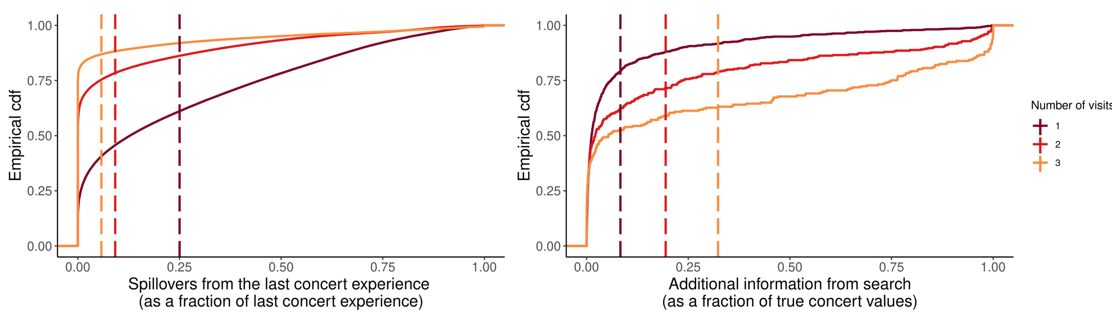


Figure 8: Posterior means of prior information as a fraction of true concert values: 2 normal mixture components

Figure 9(a) illustrates the degrees of learning spillovers at each visit. To create the empirical cdf, I take the actual observed concert choices by sample customers and their similarities to other upcoming concerts, and use them to convert the structural parameters (ρ) to the fraction of the previous concert experience that is carried over (g_{ikt} in Equation (28)). For example, 0.25 on the x-axis indicates that the predicted concert value of an upcoming concert is updated by $0.25 \times \Delta_{ijt}$(information discrepancy realized after visiting concert j). Each line represents how learning spillovers takes place at different visits. Two patterns are noticeable. First, the average learning spillover from the initial concert value to all the upcoming concert values is 0.25. In other words, on average, if a customer comes to the symphony center for the first time with a prior belief of zero concert value, 25% of the concert value experienced at the first visit is added to the prediction of an upcoming concert value. This suggests a significant impact of a visitor's initial concert experience on her perception of all the upcoming concerts. Second, experience spillovers take place more locally as the number of visits goes up (light orange line). This implies that the same set of concerts can generate different consumer perceptions of the concert value distribution if those concerts are experienced in different orders.



(a) Learning from consumption

(b) Learning from search

Figure 9: Learning spillovers from previous concert visits and additional information acquired from search at different visits

Figure 9(b) shows search intensity $\phi_{it} \in [0, 1]$ across visits. Here, I take the actual observed concert choices by sample customers and use them to convert the structural parameters (ϕ) to the fraction of the true convert values that is added to customer beliefs (ϕ_{it} in Equation (31) and (34)). For example, 0.25 on the x-axis indicates that the predicted concert value of an upcoming concert is updated by $0.25 \times \widehat{Q}^*_k$(true concert value of an upcoming concert k). Each line represents how the amount of information acquired from search varies at different visits. The plot suggests that learning from search shows the opposite pattern to that of learning from previous concert experiences. First, the average search intensity is low (below 10%) after the first visit; in other words, less than 10% of the true concert values are recovered via search. However, the amount of search increases over visits so that more than 30% of the true concert values are recovered via search ($\phi_{it} > 0.30$) after 3 visits on average.

Low prior information about the underlying concert values, high experience spillovers, and low adjustment via search after the initial visit jointly justify early customer attrition at the symphony center level. Given the lack of prior information about the concert values at the purchase stage (Figure 8), new consumers who randomly purchase low-value concerts in their first visits would only generalize the information (Figure 9(a)) without adjusting it via search (Figure 9(b)), which results in churn at the symphony center level.

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

9 Counterfactuals

Two sets of counterfactuals are run with different goals. The first set studies how to effectively reverse churn, i.e., how to re-attract consumers who have already exposed to the low-value concerts in their first visits. The second counterfactual looks at the trade-offs between increasing and decreasing concert variety given the information problem at the purchase stage.

9.1 Reversing churn

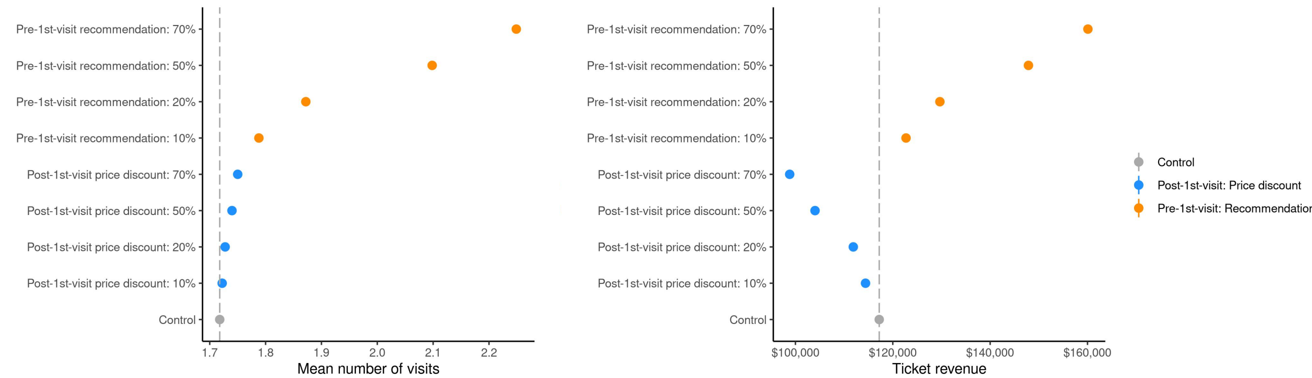
Here, a control group consists of consumers who experienced low-value concerts in their first visits. The objective is to re-attract these consumers, i.e., raising the average number of return visits of this specific group of consumers via marketing interventions.

Table 9: Reversing churn: Counterfactual design

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

Among 10000 sample customers I estimate the structural parameters for, I form a control group by selecting those customers who initially visited low-value concerts in the data. I sample those whose first concert values are lower than 10th-percentile of the concert value distribution, which results in 1509 customers. Two types of marketing interventions are applied to the group: 1) price promotion on the second visit, and 2) a probabilistic increase in their experienced concert values at the first visit to be the median concert value (e.g., via recommendation system or choice architecture). The second intervention, which takes place before the first visit and lowers the chance of initial exposure to low-value concerts, is presented to compare the benefit of marketing intervention before and after the visits to the low-value concerts. Each type of interventions has four levels of treatment, which is summarized in Table 9. Under different interventions, purchase sequence for each customer is simulated and the total number of visits is averaged across the customers over 100 weeks. 100 simulations are made for each levels of intervention.

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



(a) Number of visits

(b) Ticket revenue

Figure 10: Counterfactual results: Effectiveness of different marketing interventions on reversing churn

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

9.2 Trade-offs between increasing and decreasing product variety

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

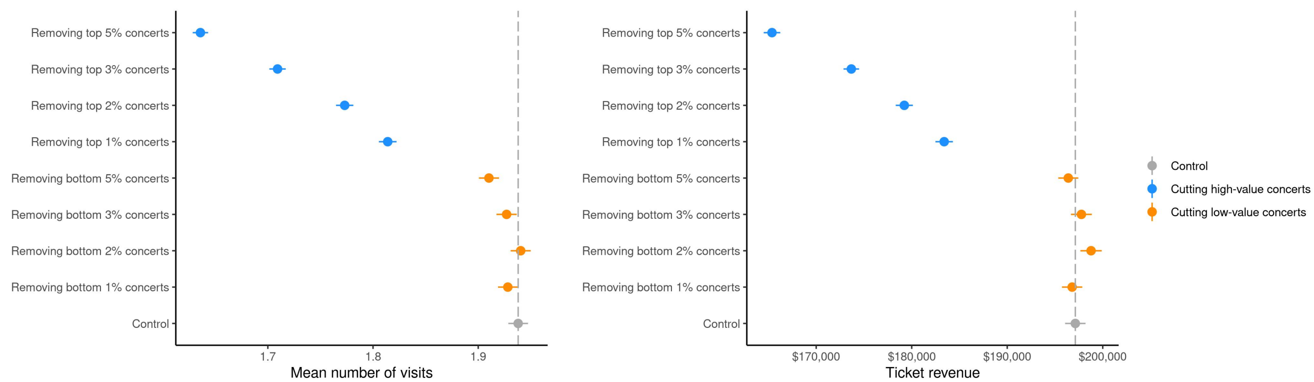
To investigate the magnitude of these trade-offs, I simulate consumer purchases for 100 weeks when some of the concert offerings are removed. Table 10 describes the counterfactual design.

Figure 11 summarizes the results. First, average number of visits drops as the number of variety

Table 10: Trade-offs in product variety: Counterfactual design

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

goes down in most conditions because of the lower arrival rate. However, when low-value concerts are removed, the mean number of visits only drops slightly or even stays statistically the same with the number in control condition, suggesting the positive effect of reduced variety offsetting the negative effect. Second, the simulated ticket revenue is even higher when the bottom 2% concerts are dropped from the product offering. Since the reported ticket revenue does not take into account cost savings from staging less concerts, the net profit from concert offering reduction would be larger and positive.



(a) Number of visits **(b)** Ticket revenue

Figure 11: Counterfactual results: Reducing concert variety before the first visit

Figure 12 shows the predicted net profit of reducing concert variety as a function of average concert production cost. To create the plot, I first compute a scale factor that matches the simulated ticket revenue in a control group to the observed market size, and multiply the simulated ticket revenues in treatment conditions by the same scale factor. Change in net profit is computed with

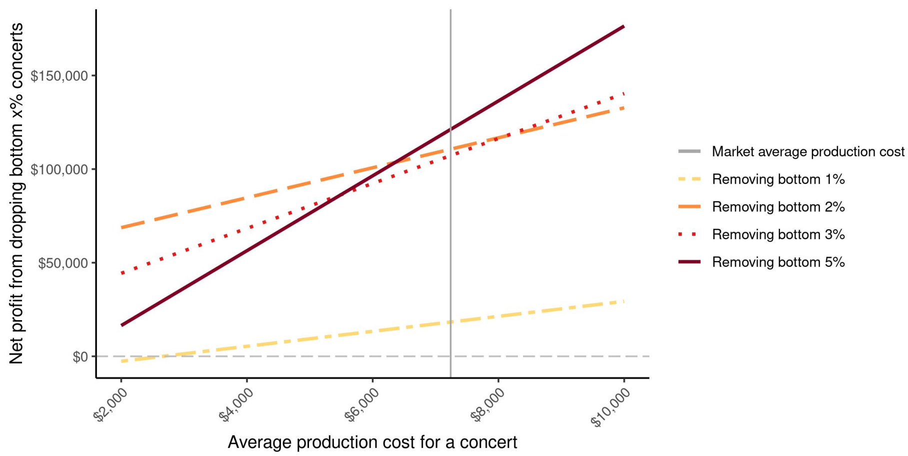


Figure 12: Predicted net profit from dropping low-value concerts as a function of average production cost

the following equation:

$$\begin{aligned}
&\Delta\text{Net profit} = \Delta\text{Ticket revenue} - \Delta\text{Production cost} \\
=&(\text{Scaled ticket revenue in treatment condition} - \text{Scaled ticket revenue in control condition}) \\
&- (\#\text{Concerts in treatment condition} - \#\text{Concerts in control condition}) \cdot (\text{Av. production cost}).
\end{aligned} \tag{40}$$

To remove 1%, 2%, 3%, and 5% of the concerts in 100 weeks means to drop 4, 8, 12, and 20 concerts respectively. I calculate the change in net profit when the average production cost per concert ranges from $2,000 to $10,000. The graph indicates that removing bottom 5% concerts is predicted to create almost $120,000 of incremental profit from the entering cohort for 100 weeks if I assume the production cost to be at the reported market average.[15] However, removing bottom 5% of the concerts will give less incremental profit than removing bottom 2% if the average production cost per concert is lower than $6,000, which highlights the tension in determining the optimal level of product variety.

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

10 Conclusion

This paper shows how incomplete information can lead churn among new customers in two steps. First, the paper shows a pattern inconsistent with what a model with fully informed preferences would predict, demonstrating incomplete information among first-time customers. Second, it documents a significant impact of the estimated consumption experience at the very first visit on subsequent churn. These two pieces of evidence explain the mechanism behind why customers churn at the firm level after a single product trial; customers who buy low-value products (due to the incomplete information about the underlying consumption utilities at the purchase stage) generalize their initial experiences to all the other untried products and leave the firm (due to the incomplete information about the range of available consumption utilities offered by a given firm).

Counterfactual analyses suggest that it is challenging to earn back first-time customers ex-post after they are already exposed to low-utility experiences in their initial consumption stage. While different theories and heuristics imply the potential significance of the initial consumption experience in forming subsequent actions of consumers, little empirical studies on churn have looked at the lasting effect of first contact on the following customer relationship management. Also, little is known about why such a high number of churn events are observed at the very early consumption stage. The paper fills in this gap by first estimating the true consumption utilities and using it to identify its causal impact on churn under imperfect information.

This paper opens a broad discussion on how consumers learn about the value of a firm based on only a few samples of experiences, although a firm represents a collection of diverse experiences instead of a small number of standardized products. Future research may extend our understanding of how imperfect information and consumer learning affects firms' optimal marketing strategies to prevent churn. For example, understanding how capacity constraints (e.g., only 100 seats available for each show) affect optimal product variety or pricing policy given the risk of information-induced churn might be an important issue for a firm to solve.

References

Anderson, E., Lin, S., Simester, D., & Tucker, C. (2015, oct). Harbingers of failure. *Journal of Marketing Research*, *52*(5), 580–592. doi: 10.1509/jmr.13.0415

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

Bauer, E. (2017). All about customer acquisition cost (cac). *Propeller*.

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

Cuffe, J., Narayan, A., Narayanan, C., Wadhar, H., & Wang, J. (2018). Retention. *Sequoia Capital Publication*.

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

Dew, R., Ansari, A., & Li, Y. (2020). Modeling dynamic heterogeneity using gaussian processes. *Journal of Marketing Research*, *57*(1), 55-77. Retrieved from `https://doi.org/10.1177/0022243719874047` doi: 10.1177/0022243719874047

Doraszelski, U., Lewis, G., & Pakes, A. (2018, mar). Just starting out: Learning and equilibrium in a new market. *American Economic Review*, *108*(3), 565–615. doi: 10.1257/aer.20160177

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. (2007, jan). How to project customer retention. *Journal of Interactive Marketing*, *21*(1), 76–90. doi: 10.1002/dir.20074

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

Gessner, K. (2018). Takeout takeover: Uber eats now bigger than grubhub in 15 major u.s. cities. *Bloomberg Second Measure*.

Gessner, K. (2019). Ahead of ipo, airbnb's consumer sales surpass most hotel brands. *Bloomberg Second Measure*.

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.*

Huang, Y., Ellickson, P. B., & Lovett, M. (2018). Learning to set prices in the washington state liquor market. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3267701

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

Narayanan, S., & Manchanda, P. (2009, may). Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, *28*(3), 424–441. doi: 10.1287/mksc.1080.0410

Narayanan, S., Manchanda, P., & Chintagunta, P. K. (2005, aug). Temporal differences in the role of marketing communication in new product categories. *Journal of Marketing Research*, *42*(3), 278–290. doi: 10.1509/jmkr.2005.42.3.278

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

Simester, D. I., Tucker, C. E., & Yang, C. (2019, sep). The surprising breadth of harbingers of failure. *Journal of Marketing Research*, *56*(6), 1034–1049. doi: 10.1177/0022243719867935

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

Tan, G. (2016, June). Blue apron: Inside the box. *Bloomberg Second Measure*.

Trivedi, N. B. (2017). How blue apron compares to other subscriptions in one graph. *Medium*.

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

Weitzman, M. L. (1979, may). Optimal search for the best alternative. *Econometrica*, *47*(3), 641. doi: 10.2307/1910412

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

Zhao, Y., Zhao, Y., & Helsen, K. (2011, apr). Consumer learning in a turbulent market environment: Modeling consumer choice dynamics after a product-harm crisis. *Journal of Marketing Research*, *48*(2), 255–267. doi: 10.1509/jmkr.48.2.255

Appendix

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

Inclusion of genre fixed effects in Equation 22 allows that genre preferences are already known to consumers instead of being learned. More specifically, constructing a concert value measure as a score within a genre has two implications. First, it allows that there can be systematic differences in tastes for genres between experienced and inexperienced visitors, as the measure is net of experienced customers' baseline preferences for different genres. Second, it assumes that customers have full information about their preferences for each genre, and that there is a component to be learned for each concert within a genre. For example, all chamber music concerts can be ranked relative to one another based on their average match value to the market (which is to be learned), but the ranking is created separately for the chamber music concerts and for Jazz concerts. The choice of genre is fully driven by consumer tastes that are known to consumers from the beginning.

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

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

To check the validity of the estimated concert values, I compare the concert value estimates with the popularity of featured artists, composers, and pieces performed using the Billboard charts data. I scrape the information on top 20 classical music albums in Billboard's weekly chart during the data period and create a Billboard score for artists, composers, and musical pieces based on the their frequency of appearances on the charts. [16]. Next, for each concert I create a concert-specific Billboard index by summing the Billboard scores if any of the composers, artists, or pieces on Billboard charts are staged in a particular concert. Figure 13 shows that the estimated concert values are positively correlated with such billboard scores, especially among those concerts with high estimated concert values.

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

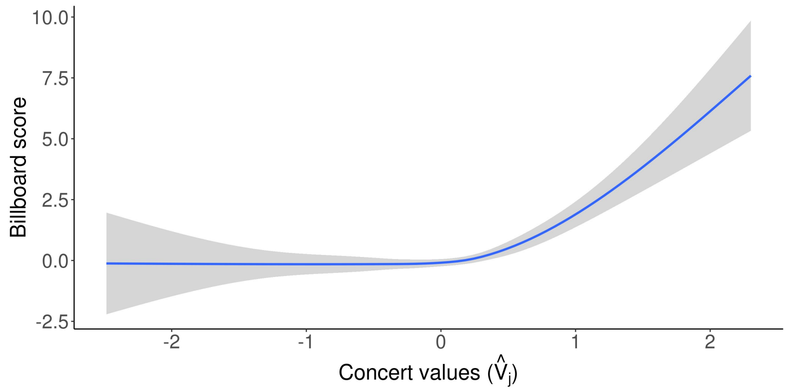


Figure 13: Correlation between the estimated concert values $\widehat{Q}^*_j$ and Billboard rankings

For further validity check, I run the following regression:

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

If the estimated concert values are valid, it should have a 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 a positive correlation with the estimated product values. Table 11 shows that all coefficients are positive as predicted. I only include non-bundle ticket purchases in this regression, but the result stays the same when I also include bundle purchases.

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

Table 11: Correlation between the estimated concert values and other variables in the individual ticket purchase data

	Dependent variable:
	$\widehat{Q}^*_j$
log(Days of wait + 1)	0.073***
	(0.002)
log(Price paid)	0.048***
	(0.003)
$\mathbb{1}${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

used here is not affected by it. Table 12 shows the list of top ranked concerts according to the concert value estimates. The list suggests that the concert value measure is neither in favor of niche tastes nor biased against the concerts targeting the mass public.

Table 12: Top ranked concerts according to the estimated concert value measure

Rank	Performance name	Category name
1	Verdi	Specials
2	Oberion Trio	Chamber
3	Dudamel/Yo-Yo Ma	Specials
4	Alexandre Tharaud	Guest Piano
5	Rachmaninov 3	Main
6	Gershwin	Specials
7	Youth in Music	Emerging professionals, fusion
8	Simon Bolivar	Specials
9	2015 Festival	Emerging professionals, fusion
10	Mendelssohn Elijah	Specials
11	Kodo	Non-western
12	LOUIS	Specials
13	Lincoln Bicentennia	Specials
14	Silk Road	Non-western
15	Viva Brazil: Ma	Non-western
16	Silk Road Ensemble	Non-western
17	Movies: Williams	Movies
18	Big Green Meadow	Family
19	Brass quintet	Specials
20	P.D.Q. Bach	Specials

A.3 Incorporating heterogeneity into the estimator of concert values - ONLINE APPENDIX

The estimator can be easily modified to explicitly allow for heterogeneity in perceived product values. One way to incorporate heterogeneity is to first cluster experienced consumers based on their purchases and to construct a separate product value measure for each customer group. Table 13 illustrates the approach.

Table 13: Clustering experienced consumers based on purchase decisions

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

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

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

Let $Y = (Y_1, \dots, Y_N)$ denote a binary vector of concert purchases by N experienced customers where $Y_i = (y_{i1}, \dots, y_{iJ})$ and J is the total number of concert offerings. Each customer belongs to one of G classes and each class represents different tastes for concerts. There are two main sets of parameters: probability that an individual belongs to a group $g \in \{1, \dots, G\}$ (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 $G_i = (c_{i1}, \dots, c_{iG})$ is a binary vector that represents i's true group membership; $c_{ig} = 1$ if i's membership is $g \in \{1, \dots, G\}$ and 0 otherwise. If G_i is observed with Y_i, I can write the likelihood of (Y_i, G_i) to be

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

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

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

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

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

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

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

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

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

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

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

$\widehat{\theta}$ and $\widehat{\pi}$ for 2 latent groups ($G = 2$) estimated via this algorithm are presented in Figure 14.

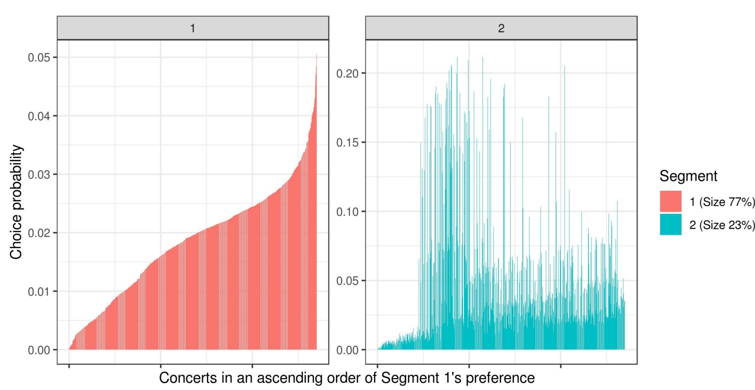


Figure 14: Estimated concert choice probabilities of experienced customers by latent group

Figure 14 shows the estimated choice probabilities of individual concerts of two latent customer groups. The x-axis denotes individual concert in an ascending order of Group 1's preference, and the y-axis represents the choice probability of each concert by group. Two patterns are noticeable. First, there is a big asymmetry in size between Group 1 and Group 2, which implies that the majority of consumers are clustered as the same latent class. Second, both groups agree on the concerts they do not prefer according to the low choice probabilities of the concerts located on the left side of the x-axis.

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

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

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

$$Churn_{i\tau} = b_{q1}\widehat{Q}_{i\tau} + b_{q2}\widehat{Q}_{i\tau-1}b_{q3}\widehat{Q}_{i\tau-2} + Z_{i\tau}\Gamma + e_{i\tau} \tag{A.0.2}$$

where $Churn_{i\tau} = 1$ if consumer i churns after τth visit and $Z_{i\tau}$ has three more sets of control variables in addition to those in Table 6).

$$\begin{aligned} Z_{i\tau} = \{ & \log(\text{Seat Quantity}_{i\tau}), \log(\#\text{ Days purchased in advance}_{i\tau}), \log(\text{Price paid}_{i\tau}), \\ & \text{Concert hour}_{i\tau}, \text{Concert day of week}_{i\tau}, \text{Concert month}_{i\tau}, \text{Concert genre}_{i\tau}, \\ & \text{Popularity of the concert among the first-time customers}_{i\tau}, \text{Zip code}_{i}, \\ & \text{Seat Area}_{i\tau}, \text{Promotion Type}_{i\tau}, \text{Ticket Sales Channel}_{i\tau}\}. \end{aligned} \tag{A.0.3}$$

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

Figure 15 plots the smoothed conditional mean of the residual probability of churn ($e_{i\tau}$ in Equation (A.0.2)) with respect to the initial concert values experienced. Even after controlling for a richer set of confounders, the negative correlation stays robust, which further corroborates the

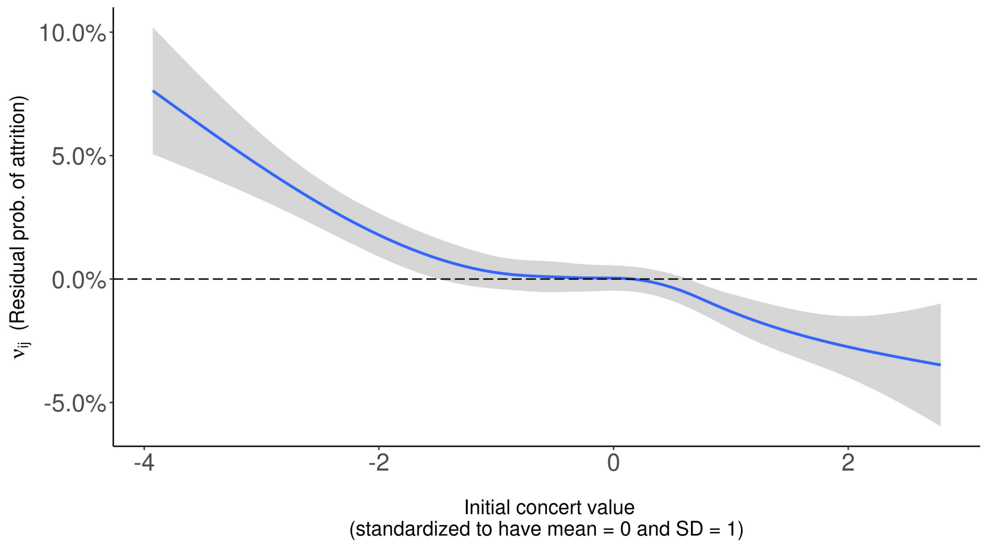


Figure 15: Experienced concert value at the first visit vs. Residual probability of churn after the first visit (Linear probability model)

impact of experienced concert value on subsequent churn decision.

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

To see if there is any additional effect of initial concert experience on customer lifetime value, I run the same regression as in Equation (23) using two alternative outcome variables: conversion to a regular customer (with 10+ visits) and conversion to a donor.[17] Table 14 summarizes the result.

Coefficients on the experienced concert value ($\widehat{Q}_{i\tau}$) are significant and positive when the outcome variable is whether a customer stays for more than 10 visits (Column 1 to 3). As in Table 7, the effect of the most recent concert value persists even when the previous concert value is controlled for. Initial consumption experience also shows a significant and positive correlation with conversion to a donor (Column 4 to 6). However, the correlation becomes insignificant for visitors after three visits.

Table 14 also reports the predicted changes in the outcome variables under 1-standard-deviation increase in most recently experienced concert value ($\widehat{Q}_{i\tau}$).[18] 1-standard-deviation increase in initially experienced concert value leads to $253K increase in ticket revenues over 7 years for a given consumer cohort. This estimate is much larger than the predicted increase in ticket revenue when assuming no change in churn rates beyond the second visit (Table **??**).

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.

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

[18] To monetize contribution that changes in initial concert visits can make, I use the following equation to compute

Table 14: 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.

Moreover, it also has an impact on customers' subsequent donation behavior. The results implies that overall effect of initial concert experience on customer lifetime value would be large enough to make significant contribution to the symphony center's profit.

A.6 Computing concert similarities

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

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

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

Figure 16 shows the distribution of concert similarities created with Gower's metric. Concert similarities are normalized such that they lie between 0 and 1.

---

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

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

Using median total ticket price or donation generates predictions that are about 60% of the reported predictions in Table 14.

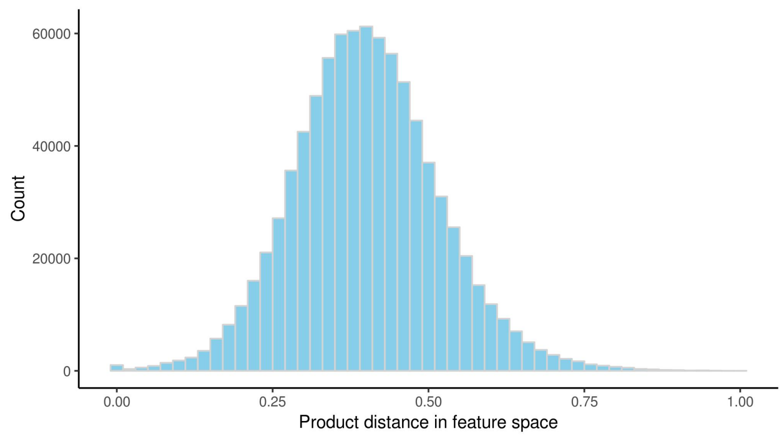


Figure 16: Distribution of concert similarities created with Gower's metric (normalized to be between 0 and 1)

A.7 Correlation between concert prices and concert values ($\widehat{Q}^*$)

Figure 17 shows the correlation between the list price and the estimated concert values within genre. As the plot shows, the correlation between concert values and the list prices is not positive in many cases or not strong even if it is positive. Some genres show no correlation (e.g., Emerging professionals) because every concert in this specific category is offered at the same price.

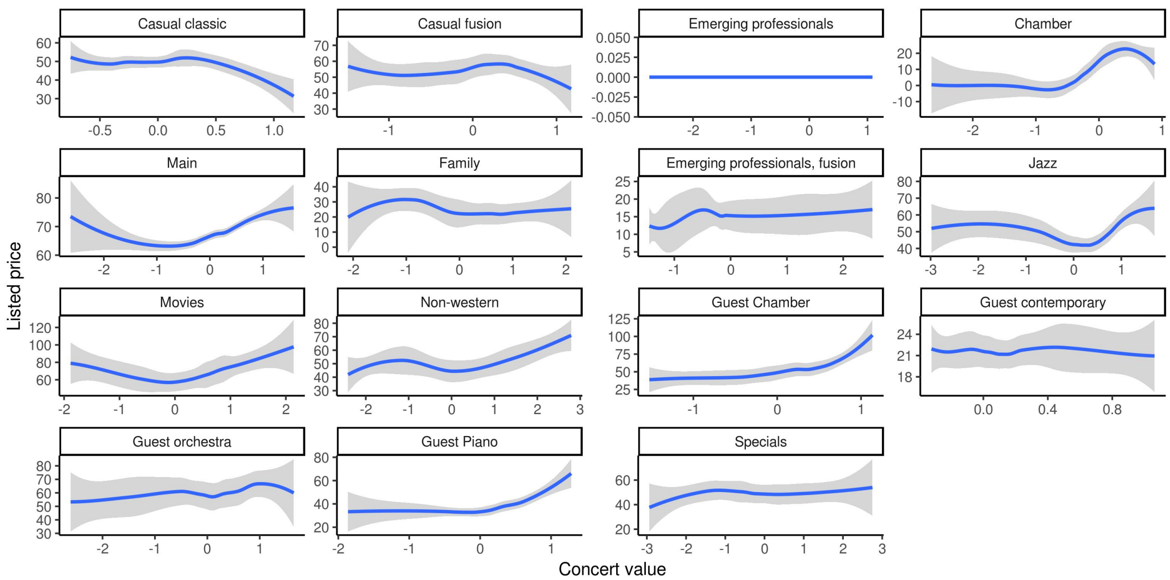


Figure 17: Estimated concert values vs. listed price