Can Willingness to Pay be Identified without Price Variation?

What Usage Tracking Data Can (and Cannot) Tell Us

Cheng Chou Vineet Kumar*

February 2021

Abstract

We study how to obtain the distribution of willingness to pay (WTP) for subscription products, where consumers pay a fixed price each period. In the absence of price variation, we demonstrate how variation in usage and subscription choice together can identify the WTP distribution and elasticities, which are primitives of interest in demand estimation. We then propose a novel estimation strategy to recover the WTP distribution when price does not vary in data. Our approach does not assume specific distributions for the unobservables. In addition, we show how the existence of price variation can help identify the functional form by which usage impacts WTP. We illustrate our method with an application to a music streaming service.

1 Introduction

Our paper studies how to obtain the distribution of consumers willingness to pay (WTP) for subscription products in the absence of price variation. Estimating the distribution of WTP, given consumer and product characteristics is an essential and the most challenging step to understand and predict demand responses, to identify how consumers value various features of the product, and to decide how alternative products should be priced. Consider the example of Netflix, which has a monthly Standard plan priced at $12.99 in the US. When the firm is interested in evaluating how demand might vary with price increases, we would need to obtain the WTP distribution so that we can infer the percentage of consumers who are willing to pay more than the new price.

Subscriptions are becoming increasingly popular across the world for both physical and digital products and services with growth over 100% in 2013–2018 (Columbus, 2018; Chen, Fenyo, Yang,

*Cheng Chou is an Assistant Professor of Economics at School of Business, University of Leicester. e-mail: cchou@le.ac.uk. Vineet Kumar is an Assistant Professor of Marketing at School of Management, Yale University. e-mail: vineet.kumar@yale.edu. We would like to thank Jesse Matheson, Ruoyao Shi and Adam Smith for feedback. We thank participants at the University of Chicago celebratory symposium of Prof. Vithala Rao, and especially Yesim Orhun, the discussant. We would also like to thank the participants of the FORMS conference at UT Dallas and Michael Braun, the discussant. We also thank participants in seminars at University of Sheffield, UCL, University of Toronto, Washington University at St. Louis, and Yale University.

and Zhang, 2018). Subscription plans are prevalent across a wide variety of industries, ranging from media to software-as-a-service to eCommerce and transportation, as detailed in Table 1.[1]

In most subscription markets, price variation is fairly rare or non-existent, except for free trials.[2] The absence of price variation presents a major challenge in identifying the distribution of WTP—how would you predict the demand response to the change of price, when price does not change at all in data. The feasibility of demand estimation in economics and marketing has depended on the presence of data with price variation. The lack of price variation poses a challenge for using the common revealed preference approach to recover the distribution of WTP, which relies on price variation, a feature common to the entire literature (Guadagni and Little, 1983; Train and Weeks, 2005; Danthurebandara, Yu, and Vandebroek, 2011; Lewbel, McFadden, and Linton, 2011). Firms in such markets set prices based on market research typically using conjoint analysis or similar survey elicitation responses (Green and Rao, 1971; Green and Srinivasan, 1978). While conjoint analysis is a very useful tool to obtain relative preferences, it has difficulty in accurately capturing valuations that are further away from the market price, and consumers have been found to have a different WTP when making actual purchase choices.

When prices do not vary, the key insight of this paper is to recognize there may be variation in other elements of the data that can be leveraged to obtain the distribution of WTP. Specifically, our approach uses variation of usage (or consumption) of the subscription product. Below we will use "usage" and "consumption" interchangeably. Overall, the typical data available in subscription settings include product usage data, subscription/churn choices, and often rich data on a variety of consumer and product characteristics. Our framework was developed to utilize these typical data.

Given this background, we examine the following research questions. First, in a subscription market setting with usage variation but *without price variation*, what can we infer about the distribution of consumer valuations from usage data? Second, in settings with price variation in addition to usage variation, what additional inference is possible?

The main contribution of this paper is to propose a novel method to identify and estimate semiparametrically the conditional distribution of WTP given product features and customer characteristics *when price variation is absent*. The semiparametric aspect of our method is that we do not assume that the unobserved heterogeneity in WTP follows any specific distribution. To the best of our knowledge, there is no research that demonstrates how to obtain the WTP distribution in the absence of price variation. Our approach does not require the presence of multiple plans and the cross-sectional inter-plan variation induced by that, e.g. Netflix has plans at \$8.99 (basic), \$12.99 (standard) and \$15.99 (premium) per month. Rather, it works with only one plan present, e.g. Apple Music only has one plan at \$9.99 per month. Our framework is quite general and is

[1]There are a number of reasons for this popularity, including low marginal costs (relative to fixed costs), reduced consumer risk, no transaction costs from the consumers' perspective, and predictability in revenue stream as well as increased loyalty from the firms' perspective (Xie and Shugan, 2001).

[2]As an example, Spotify has always set the monthly price for unlimited ad-free streaming around \$10 from 2011 to the present.

Table 1: Subscription Plans

Industry	Product or Service	Price ($)	Period	Total subscribers
Media & Entertainment	Netflix	12.99	Monthly	23 million (US)
	Spotify	9.99	Monthly	70 million (World)
	New York Times	3.75	Weekly	4 million (US)
	MoviePass	19.95	Monthly	2 million
	Kindle Unlimited	9.99	Monthly	–
	Apple News	9.99	Monthly	36 million
Software-as-a-Service	Microsoft Office 365	9.99	Monthly	120 million
	Adobe Creative Cloud (One App)	20.99	Monthly	15 million
	Dropbox Premium	9.99	Monthly	>11 million
Membership Clubs	Costco (Basic)*	60	Annual	94 million
	Amazon Prime	119	Annual	90 million
	24 hour fitness (Gym)	40	Monthly	4 million
eCommerce	Harry's	35	Monthly	–
	Birchbox	15	Monthly	2 million
	Rent the Runway	159	Monthly	6 million
Transportation	Public Transit Pass (MTA)	121	30-days	–
	Uber Ride Pass*	14.99	Monthly	–
	Jetblue "All You can Jet" Pass	699	Monthly	–

Note: Data collected Nov 2019. "–" indicates public data was unavailable.

consistent with a number of models providing microfoundations for usage, and does not restrict such usage models.

Second, we show that given the presence of usage variation, price variation is still necessary to answer additional questions (e.g. how to explain the cases where we observe subscribers with zero usage). In particular, we show how price variation (even limited price variation like only two distinct prices for the same subscription product) and usage variation together can permit the identification and estimation of a more general class of models. For example, we show that with usage variation and two distinct prices, we can write a model of WTP that can accommodate the cases where some subscribers have zero usage. Broadly, our findings point to both the potential and limitation of usage data in recovering the WTP distribution when price variation is absent.

The intuition for our main result is the following: A consumer's WTP for a subscription plan can be decomposed into (a) her expected usage of the plan and (b) her average WTP for one unit of usage. To know the distribution of the WTP for a subscribed service, we need to know the distribution of both the expected usage and the WTP per unit usage, as well as the *correlation between usage and WTP per unit usage*. We can identify the distribution of expected usage for subscribers based on the data. The identification of the distribution of WTP per unit of usage relies on the following argument. Even though price is identical among consumers, there is still variation in the "price" of *per unit* usage when there is variation of usage among consumers. We then can recover the distribution of the WTP per unit usage from the variation of the "price" of per unit usage along with either temporal or cross-sectional variation in subscription choice. Under a set of reasonable assumptions, we also address the issue of correlation between usage and WTP per unit usage.

We take our method to data using an application of music streaming, featuring monthly subscription and usage choices. We estimate the conditional distribution (on demographic characteristics) of WTP and elasticities of the WTP for its monthly streaming plan. We find that the age elasticity of usage is negative, whereas the elasticity of the WTP with age is positive, indicating that older users use the product less, but value it more than younger users. We find male subscribers have higher WTP for the service. We also estimate the mean of log WTP for different age and gender groups.

We note that the paper has a scope beyond subscription markets in identifying WTP. The crucial aspect is that we need a separation of purchase and consumption and data on both. We discuss in Section 7 how it can be applied to more broadly, for instance, to packaged goods.

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 details the identification of the conditional distribution of the WTP and the conditional mean of the log WTP, given product and consumer characteristics, which can be used to derive the demand curve and the elasticities of the WTP. Section 4 studies the role of price variation in identifying the WTP distribution. Section 5 examines the small sample properties of the estimator. Section 6 uses the approach in an empirical application of music streaming to demonstrate its potential value. Section 7 concludes the paper. The appendices contain technical proofs.

2 Literature

There are multiple streams of literature focused on measuring and characterizing WTP or the distribution of valuations. See Breidert (2007) for a comprehensive overview. There are a few different approaches to eliciting WTP, either at an individual level or in obtaining a market-level aggregate. An important distinction should be made between methods that use stated preference to obtain *hypothetical* WTP, and that use revealed preference to obtain *real* WTP. In the real WTP case that involves consumer choice, to the best of our knowledge, there is currently no general method that can obtain the WTP distribution in the absence of price variation. On this point lies the primary contribution of this paper.

Within the *stated preference* stream of literature, customer populations are surveyed to obtain an estimate of WTP. Such approaches are typically used to obtain hypothetical WTP since consumers do not have to actually pay a price or face financial consequences. These include direct surveys of consumers or buyers, which remain used in contingent valuation type settings without product variation. For example, consumers might be asked how much they value particular public and environmental goods, e.g. a park (Mitchell and Carson, 2013; Hanemann, 1994). The appeal of this methodology is in its simplicity and in obtaining an economically relevant quantity, although researchers have long pointed out the challenges in obtaining an accurate estimate (Diamond and Hausman, 1994; Hausman, 2012).

The conjoint analysis method developed in marketing and has a strong stream of research (Green and Rao, 1971; Green and Srinivasan, 1978). See Rao (2014) for a comprehensive perspective, and Ding (2007) for incentive compatible conjoint. Conjoint guides consumers into making rank-ordered preferences from a limited choice set. With sufficient observations, it is possible to obtain an individual level willingness to pay not just for the overall good, but for its constituent features, e.g. battery life in a device.

There are many advantages of this stated-preference approach. First, it is relatively easy to implement, and allows for exogenous variation in product characteristics, prices and choice sets available to consumers. These sources of exogenous variation are powerful in providing clear identification through induced variation. Another advantage is that it can be used to test how the market values hypothetical improvements or changes in advance of actually making them. Within this stream there are two broad approaches: direct surveys and choice-based conjoint. Direct surveys ask individuals to place a monetary value on a product or service (contingent valuation). Conjoint on the other hand asks consumers to rank order choices, which can vary based on price as well as other characteristics. A disadvantage of conjoint is that it is usually set at market prices, implying that with higher dispersion of willingness to pay, valuations that are further away from the market price will not be accurately captured. Moreover, there is the key question of whether stated preferences correlate with actual behavior or revealed preferences. Several studies have found significant differences in elicited valuation depending on the specifics of the method used to obtain it. As detailed across a variety of studies, the stated or hypothetical WTP is often found to be higher than revealed WTP (Kalish and Nelson, 1991; Wertenbroch and Skiera, 2002; Voelckner, 2006).

Next is the well-established literature on demand estimation *using observational data*, either at the individual consumer level like in much of the marketing literature (Guadagni and Little, 1983), or market-level like in (Berry, 1994; Berry, Levinsohn, and Pakes, 1995) and related literature. In all of these cases, the idea of price variation is central to identification (see for example Rosen 1974; Heckman, Matzkin, and Nesheim 2010; Shi 2019). In addition, endogeneity is often an important concern in demand estimation (Villas-Boas and Winer, 1999). The typical issue is that prices are set based on unobservable characteristics of products or based on market or consumer characteristics. Thus, we cannot rely on exogenous variation to be able to identify demand. Researchers typically rely on instrumental variables or control function approach to identify demand and WTP. While this problem has been well recognized, it is not just a theoretical concern. Since the time of Trajtenberg (1989), it has been noted that without carefully accounting for unobserved characteristics, positive price coefficients can be obtained, implying consumers prefer to pay more, all else equal.

Within marketing, there is a rich stream of literature focusing on specifying and estimating rich models of consumers heterogeneity. These models either use random coefficients for individual households or a hierarchical Bayesian approach, and help in designing and evaluating targeted interventions to specific households (Rossi, McCulloch, and Allenby, 1996). This focus on individual heterogeneity is very helpful in targeting promotions (e.g. coupons) at the individual or household level, allowing more efficient generation and capture of surplus by the firm. The present paper shares many features with this stream in the sense that we are interested in characterizing the valuation distribution at the individual customer level, and potentially condition it on observable demographic characteristics.

Another set of papers involve field experiments, where researchers have carried out different experimental designs to elicit the WTP distribution. These involve either an auction based approach (Vickrey auction) as in Noussair, Robin, and Ruffieux (2004) or involve a stochastic price generation mechanism (Becker, DeGroot, and Marschak, 1964) that induces incentive compatibility among the participants.

It is striking that *none of the above methods* provide any help when there is no price variation in the data or under an experiment. Even the use of instruments is infeasible in such a case because there is no instrument that can be correlated with a constant price, and be uncorrelated with the unobserved errors. There are a small set of papers that include demand estimation when prices are fixed. In a model with multiple products, i.e. print and online newspapers, Gentzkow (2007) uses moments derived from supply-side first order conditions to obtain identification. In contrast, our approach does not assume a supply-side model or require multiple products. However, we do need access to usage data, consistent with our focus on subscription markets.

A related paper is Nevo, Turner, and Williams (2016), who estimate demand for residential broadband using usage (download/upload in GBs) and plan choice (e.g. unlimited usage plans vs usage-based plans) data when subscribers face a three-part tariff, featuring an *overage price* for each GB of usage in excess of a specified allowance. They model a forward looking consumer as realizing that the opportunity cost of usage depends on the distance to this allowance or quota, changing their *shadow price*. Their identification strategy for demand estimation exploits the *variation of shadow price, induced by usage*, as the accumulated usage approaches the included allowance. In contrast, our identification arguments do not rely on the presence of *overage price*. This is important in practice because subscription products typically do not use three-part tariff pricing. All the examples in Table 1 and the music streaming service that we use in our empirical application have no overage charges.

3 Subscription Model

For concreteness, we consider a monthly music streaming service. We observe panel data about the binary subscription decisions and monthly usage (e.g. minutes of listening music using the service) by n number of consumers over T months. We first present the usage model for a subscriber. The conclusion there is that in equilibrium the indirect utility or WTP is proportional of one's optimal expected usage. We present it based on Cobb-Douglas utility, but show that any homothetic utility function can be used in its place. To incorporate consumers heterogeneity, we then connect the usage and valuation of the service with consumers heterogeneity. We then show the identification of the distribution of WTP among different groups of consumers as well as the estimation steps.

3.1 Usage Model

Let t denote one month. Consumer i has to solve a time allocation problem between using the service Q, and other activities, Q_0. Note that for a subscriber there are no monetary costs of usage, since the subscription cost has already been paid. This time allocation problem depends on consumer i's expected leisure time L_{it} and a demand shock ϕ_{it} at the beginning of month t. The usage demand shock ϕ_{it} shifts the utility of Q upwards or downwards in a variable manner each time period. The consumer's problem is:

$$W_{it} \equiv \max_{Q_{it}, Q_{0,it}} u(\phi_{it} Q_{it}, Z_{0,it}) \qquad \text{subject to} \qquad Q_{it} + Q_{0,it} \leq L_{it} \tag{1}$$

Our method does not require the time budget to be observable. We introduce stochasticity in the consumer's decision through time-varying ϕ_{it} and time budget L_{it}. The first shock ϕ_{it} shifts the marginal utility of using the music streaming service, and it varies over time. In some time periods, consumers might value usage more due to unobservable factors. Similarly, the other stochastic factor, L_{it} could track the amount of free leisure time that consumers have, as it varies over time.

Consider the utility function to be Cobb-Douglas with $u(\phi Q, Q_0) = (\phi Q)^a Q_0^{1-a}$.

We can solve the consumer's problem and express the optimal expected quantities as:

$$Q^*_{it} = a L_{it} \qquad \text{and} \qquad Q^*_{0,it} = (1-a) L_{it} \tag{2}$$

We can now represent the indirect utility as:

$$W_{it} = Q^*_{it}(a, L_{it}) \underbrace{\left[\phi^a_{it} \left(\frac{1-a}{a} \right)^{1-a} \right]}_{\text{denoted by } \alpha_{it}} \tag{3}$$

Thus, we see above that the model of microfoundations can express indirect utility as proportional to the *optimal* or equilibrium consumption quantity and a factor that depends on model primitives. This "proportional" relationship only holds when consumers choose optimal usage. We assume that consumer i's subscription decision rule is

$$S_{it} = \mathbb{I}(W_{it} > P) = \mathbb{I}(\ln W_{it} > \ln P).$$

Implicitly, we normalize the expected (indirect) utility of not subscribing to be zero.

We observe that this above formulation does not only apply to Cobb-Douglas but more generally according to Proposition 1 (All proofs are in the Appendix). We can use any homothetic direct utility function (including perfect substitutes, perfect complements, Leontief, Cobb-Douglas, CES etc.).

Proposition 1. *If the direct utility function $u(\phi q, z)$ is homothetic, then the indirect utility can be represented to be multiplicative in optimal consumption as: $V = q^*(L, p_q, p_z)g(p_q, p_z)$.*

Next, we will link the term of demand shock α_{it} and expected optimal usage Q^*_{it} with consumers' heterogeneity. We let α_{it} depend on observable and unobservable characteristics as:

$$\ln \alpha_{it} = \beta' X_{1it} + U_{it},$$

where X_{1it} is a part of X_{it}. We can normalize the mean $\mathrm{E}(U_{it}) = 0$ by including an intercept term in X_{1it}. Note that we don't make any specific distributional assumptions on the unobservable U_{it}. The observables X_{1it} may include a number of factors about consumer i and/or the subscribed product that impact purchase. When some elements of the vector X_{1it} are correlated with unobserved heterogeneity U_{it} in WTP per unit usage or unobserved heterogeneity V_i in usage below, we say these variables are "endogenous". For endogenous variables, we need a vector of instrumental variables (IV) Z_{it}, for which the restrictions are detailed in Assumption 2. When X_{1it} is uncorrelated with U_{it} and V_i, $Z_{it} = X_{1it}$.

In Assumption 1, we link the expected usage Q^*_{it} with the observables X_{2it} and actual observed usage Q_{it} for current subscribers. Under this assumption, we can use fixed effect estimator to identify and estimate the part of expected usage driven by the time varying part of X_{2it} for all consumers regardless of their subscription choices. This is useful because we do not observe usage by non-subscribers, but we do observe their characteristics. Here X_{2it} is a part of X_{it}, which could have intersections with X_{1it}, but cannot be a subset of X_{1it}. The reason why we don't allow X_{2it} to be a subset of X_{1it} will be clear after stating Assumption 2.

Because we will apply fixed effect estimator to usage panel data (Q_{it}, X_{2it}), it is useful to clarify which are time varying and which are time invariant variables in X_{2it}. We also need to clarify the overlapping relationship between X_{1it} and X_{2it}, which impacts the identification results. Figure 1 illustrates this relationship and the notation used to distinguish them.

Assumption 1 (Reduced form usage model). *(i) The log of expected usage $\ln Q^*_{it}$ has the following reduced form regardless of the subscription choice S_{it},*

$$\ln Q^*_{it} = \gamma' X_{2it} + V_i = \gamma'_a X_{2i,a} + \gamma'_b X_{2i,b} + \gamma'_c X_{2it,c} + V_i, \tag{4}$$

where the unobserved fixed effect V_i can be correlated with U_{it} in the specification of WTP per unit usage (α_{it}). Assume that $\mathrm{E}(V_i) = 0$. Among X_{2it}, $X_{2it,c}$ are time varying covariates,

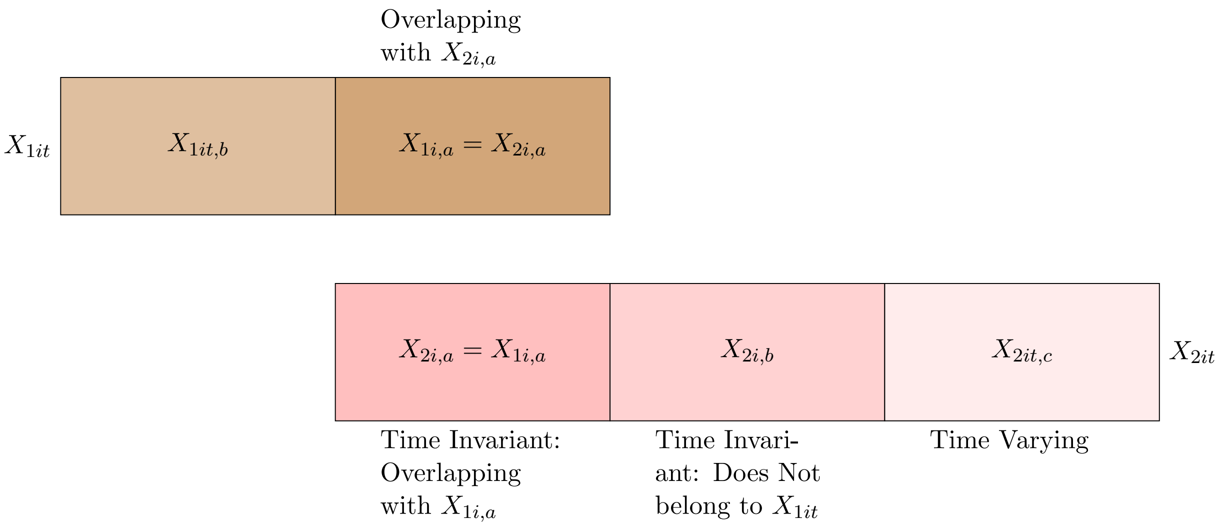


Figure 1: Relationship between X_{1it} and X_{2it}

$X_{2i,a}$ and $X_{2i,b}$ are time invariant, and $X_{2i,b}$ does not belong to X_{1it}, but $X_{2i,a}$ belongs to X_{1it}. See also Figure 1 for illustration.

(ii) The observed actual usage Q_{it} when $S_{it} = 1$ is given by:

$$\ln Q_{it} = \ln Q^*_{it} + \varepsilon_{it},$$

where ε_{it} is serially uncorrelated random shock.

(iii) Strict exogeneity: $\mathrm{E}(\varepsilon_{it} \mid X_i, U_i, V_i) = 0$, *where* $X_i = (X_{i1}, \dots, X_{iT})$ *and* $U_i = (U_{i1}, \dots, U_{iT})'$.

Assumption 1 (i) essentially separates out the impact of individual consumer fixed effects on expected usage from other observable factors. This rules out systematic time-varying unexpected unobservables that are known to the consumer, but are unobserved by the researcher. For example, the user might know they are traveling over the next week and might be using the service less. Assumption 1 (ii) restricts the expected usage to be rational expectations for the actual usage (i.e. we don't have systematic deviations). The strict exogeneity Assumption 1 (iii) can intuitively be interpreted as implying that the error in actual usage is uncorrelated with the observable characteristics X_{1it} and X_{2it} across each of the time periods. One implication of strict exogeneity is that X_{2it} cannot include the past actual usage in order to estimate γ, which is a common restriction in dynamic panel data model.

How do Consumers Form Usage Expectations? It is important to clarify that in order to satisfy the above assumption, what information is available for consumers to form their expected usage level in our model. More specifically, our model characterizes the following:

(i) Usage for past period (month) is known and observed by the consumer (and researcher) at the end of the period.

(ii) Exact usage is not known in advance of purchase (each month).

(iii) Consumers form *rational expectations* of usage for future period (month) when they decide to make a purchase. These expectations are formed based on evolution of X_{2it} variables, which impact usage.

(iv) Consumers do not "look backwards" to see what they used last month in forming usage for next month. Therefore, high past month usage by itself does not predict high next month usage. A specification like $Q^*_{i,t+1} = f(Q^*_{i,t})$ would thus be excluded.

(v) X_{2it} variables which impacted usage in period t are allowed to be persistent, so expected usage may be higher next period, since it is possible to have $X_{2i,t+1} = h(X_{2it})$. More generally, we can use a distributed lag model to capture the dependence of current expected usage on past usage by letting some of $X_{2i,t+1}$ be equal to some of $X_{2i,t}$, $X_{2i,t-1}$.

We note that perfect foresight can also be accommodated in this framework. In such a case, the consumer knows the usage exactly (so (ii) and (iii) above do not hold), which may be perhaps less realistic. However, it is worthwhile to note that two "extremes" of rational expectations and perfect foresight are both permitted in the model. Also, another way to model consumer expectations might be to consider that consumers form expectations over the evolution of X_{2it} and then obtain expected usage from $\ln Q^*_{it} = \gamma' X_{2it} + V_i$, rather than directly forming expectations over Q^*_{it}. Such an interpretation would also be broadly consistent with our framework.

Connecting WTP and Usage Let $\eta_{it} \equiv U_{it} + V_i$, and η_{it} can be interpreted as the unobserved heterogeneity in the log of WTP that cannot be explained by observed consumers and/or product characteristics X_{it} because

$$\ln W_{it} = \ln \alpha_{it} + \ln Q^*_{it} = \beta' X_{1it} + \gamma' X_{2it} + \underbrace{U_{it} + V_i}_{\eta_{it}}. \tag{5}$$

The correlation between fixed effect V_i and unobserved heterogeneity U_{it} in WTP per unit usage has an important implication: Even when X_{2it} and V_i are uncorrelated, one cannot consistently estimate γ by pooled ordinary least square (OLS) or random effect estimator using only subscribers data because the subscription decision (i.e. churn or attrition in econometrics terminology) is not random and is correlated with usage. Fortunately, it is well known that we can estimate γ_c associated with time varying covariates $X_{2it,c}$ using fixed effect estimator under the strict exogeneity condition in Assumption 1 (iii), hence we can identify $\tilde{Q}_{it} \equiv \gamma_c' X_{2it,c}$ for all consumers, since X_{it} by our assumption is always observed regardless of subscription decision. The cost is that we have to assume strict exogeneity, which prohibits us from using the past actual usage as explanatory variable of current usage.

The next assumption lists conditions, under which we can identify the conditional distribution of WTP given X_{it} and the conditional mean of log WTP. From the former, we can derive the demand function, and from the latter, we can derive the elasticities of WTP to the change of X_{it}. Conditions like Assumption 2 are commonly made in the econometrics literature about "special regressors" (e.g. Lewbel, 2000). For the sake of exposition, it is useful to define vector $\tilde{X}_{1it} \equiv (X'_{1i,a}, X'_{1it,b}, X'_{2i,b})'$, which includes entire X_{1it} and the time invariant part of X_{2it}. Since $X_{1i,a} = X_{2i,a}$, $\tilde{X}_{1it} = (X'_{2i,a}, X'_{1it,b}, X'_{2i,b})$.

Assumption 2. *(i) (Big support of usage). The support of* $\tilde{Q}_{it} \equiv \gamma'_c X_{2it,c} \mid (\tilde{X}_{1it}, Z_{it})$ *covers the support of* $\ln(P/\alpha_{it}) - V_i \mid (\tilde{X}_{1it}, Z_{it})$*;*

(ii) (Independence of usage). $X_{2it,c} \perp\!\!\!\perp (U_{it}, V_i) \mid (\tilde{X}_{1it}, Z_{it})$*;*

(iii) (Valid and relevant IV). $\mathrm{E}(Z_{it}\eta_{it}) = 0$, $\mathrm{E}(Z_{it}Z'_{it})$ *is nonsingular, and* $\operatorname{rank} \mathrm{E}(\tilde{X}_{1it}Z'_{it}) = \dim(\tilde{X}_{1it})$*.*

(iv) $\mathrm{E}(\eta_{it}) < \infty$*.*

The support condition Assumption 2 (i) can be viewed as implying that the variation in the "explainable part of (log of) usage", i.e. $\gamma'_c X_{2it,c}$, is "broad enough" to cover the variations of (log of) "minimum usage that will make the service worthwhile" ($\ln(P/\alpha_{it})$) minus the log of unexplained individual's usage fixed effect. Thus, if our data does not feature much variation in usage, then this assumption is not likely to be satisfied. Note that the condition can be interpreted at the level of the individual consumer. However, if the WTP of the consumer does not vary much over time, then the condition is more likely to hold. We demonstrate how the failure of the support condition could bias the estimation of WTP distribution and elasticities in the numerical studies in Section 5. Note that this will only be an issue if we want to implement the nonparametric estimation. If we assume that the distribution of WTP belongs a parametric family (e.g. log normal distribution), we do not need the support condition.

It is important to point out that Assumption 2 (i) and 2 (ii) imply that X_{2it} must have time varying *excluded variables* that affect usage but do not affect the demand shifter α_{it} or WTP per unit usage (hence does not belong to X_{1it}), and such excluded variables are independent of U_{it} and V_i given X_{1it}, $X_{2i,b}$, and Z_{it}. However, such excluded variables still affect WTP hence subscription choice. To see why, note that if X_{2it} is a subset of X_{1it}, the support condition of Assumption 2 (i) fails as the support of $\tilde{Q}_{it} = \gamma'_c X_{2it,c} \mid (\tilde{X}_{1it}, Z_{it})$ includes only one point. In music streaming example, such excluded variables could be factors impacting access to internet services (e.g. WiFi or 4G network coverage, quota of cellular data). The access to internet affects the usage but does not affect one's valuation of one unit of usage, say one song, after controlling variables like income. Assumption 2 (iv) is a technical condition restricting the tail of the distribution of η_{it}.

In Table 2, we detail each of the specific assumptions, with a view to providing more details on what features of the data or empirical setting are consistent with them, and which aspects might be inconsistent.

Table 2: Summary of Assumptions

A#	Interpretation	Consistent
A1(i)	Expected usage depends on observables and unobserved individual effect	Any observable factor X_{2it} that impacts usage. Note that these factors may also impact purchase, and therefore appear in X_{1it}. Also, it is possible to accommodate time period unobservables product quality by including period dummy variables.
A1(ii,iii)	Actual usage is centered around expected usage & Difference between them ("Error") is strictly exogenous	Random noise (of unknown variance) separates actual observed and expected usage. Zero variance of ε_{it} would indicate "perfect foresight" by the consumer.
A2(i)	Variation of explanable usage needs to be big enough to cover the variation of unobserved heterogeneity in WTP.	There is at least one time varying covariate that explains usage but not the WTP given usage, and this covariate has large variation in sample, relative to WTP variation.
A2(ii)	Covariates that only explain usage are independent of unobserved heterogeneity in WTP	In a gym subscription example, suppose U_{it} and V_i are one's health awareness, and the excluded variable is weather (e.g. precipitation). Then it is reasonable to claim that weather is independent of health awareness.
A2(iii)	Valid and relevant IV	When X_{1it} is exogenous, we can just let X_{1it} be the IV for itself. Otherwise, we need an IV that is uncorrelated with $\eta_{it} = U_{it} + V_i$.
A2(iv)	The mean of unobserved heterogeneity in WTP exists.	This condition holds unless unobserved heterogeneity in WTP η_{it} follows some pathological distributions.

3.2 Identification of WTP

First, observe that using a standard panel data fixed effects model, we can identify (and estimate) γ_c in the usage equation (eq. (4)). Now that γ_c is identified, we focus on the identification of β, γ_a and γ_b in Proposition 2 and more generally, the identification of distribution of consumers' WTP in Proposition 3.

Before stating the result, we rewrite the expression of $\ln \alpha_{it}$ using the decomposition $X_{1it} = (X'_{1i,a}, X'_{1it,b})'$ as follows

$$\ln \alpha_{it} = \beta' X_{1it} + U_{it} = \beta'_a X_{1i,a} + \beta'_b X_{1it,b} + U_{it}.$$

Define $\tilde{\beta} \equiv ((\beta_a + \gamma_a)', \beta'_b, \gamma'_b)'$. Apparently, when the time invariant elements of X_{2it} do not appear in X_{1it} (i.e. $X_{1i,a}$ or equivalently $X_{2i,a}$ is empty), $\tilde{\beta} = (\beta', \gamma'_b)'$.

Proposition 2 (Can Identify: Elasticities of WTP)**.** *Suppose Assumption 1 and 2 hold. We have*

$$\begin{aligned}\mathrm{E}(\ln W_{it} \mid X_{it}) &= (\beta_a + \gamma_a)' X_{1i,a} + \beta'_b X_{1it,b} + \gamma'_b X_{2i,b} + \gamma'_c X_{2it,c} + \mathrm{E}(\eta_{it} \mid X_{it}) \\ &\equiv \tilde{\beta}' \tilde{X}_{1it} + \gamma'_c X_{2it,c} + \mathrm{E}(\eta_{it} \mid X_{it}),\end{aligned}$$

where

$$\mathrm{E}(\eta_{it} \mid X_{it}) = \mathrm{E}\left[\mathrm{E}\left(\frac{S_{it} - \mathbb{I}(\ln P - \tilde{\beta}'\tilde{X}_{1it} - \tilde{Q}_{it} \leq 0)}{f_{\tilde{Q}}(\tilde{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})} \,\middle|\, \tilde{X}_{1it}, Z_{it}\right) \,\middle|\, X_{it}\right]$$

Here $f_{\tilde{Q}}(\tilde{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})$ is the probability density function (PDF) of $\tilde{Q}_{it}$ given $(\tilde{X}_{1it}, Z_{it})$.

We have the following 2-step-least-square (2SLS) formula for $\tilde{\beta}$,

$$\tilde{\beta} = [\mathrm{E}(\hat{X}_{1it}\hat{X}'_{1it})]^{-1}\,\mathrm{E}(\hat{X}_{1it}Y_{it}),$$

where $\hat{X}'_{1it} = Z'_{it}[\mathrm{E}(Z_{it}Z'_{it})]^{-1}\,\mathrm{E}(Z_{it}\tilde{X}'_{1it})$, and

$$Y_{it} = \frac{S_{it} - \mathbb{I}(\tilde{Q}_{it} - \ln P > 0)}{f_{\tilde{Q}}(\tilde{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})}.$$

Apparently, $\tilde{\beta}$ is the probability limit of the 2SLS estimator of regressing Y_{it} on $\tilde{X}_{1it}$ using instruments Z_{it}.

We now have a formula of $\mathrm{E}(\ln W_{it} \mid X_{it})$, which can be used to calculate the elasticities of WTP to any variables in X_{1it} and X_{2it}. When X_{1it} and X_{2it} are mean independent of η_{it}, the elasticities of the WTP depends only on $\tilde{\beta}$ and γ_c. Using the usage data, one can estimate γ_c by fixed effect estimation and $\tilde{\beta}$ by a 2SLS regression. In general, one also needs to estimate $\mathrm{E}(\eta_{it} \mid X_{it})$ using its expression in the above proposition.

The "dependent" variable Y_{it} can be interpreted as all factors affecting the subscription choice S_{it}, except for the observed (log of) expected usage $\tilde{Q}_{it}$. To see this, using the definition of S_{it}, we can write the numerator part of Y_{it} as follows:

$$S_{it} - \mathbb{I}(\tilde{Q}_{it} > \ln P) = \mathbb{I}[(\tilde{\beta}'\tilde{X}_{1it} + \eta) + \tilde{Q}_{it} > \ln P] - \mathbb{I}(\tilde{Q}_{it} > \ln P). \tag{6}$$

Proposition 3 (Can Identify: CDF of WTP)**.** *Suppose Assumption 1 and 2 hold. Define the conditional choice probability (CCP) function, $\pi(\tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it}) \equiv \mathrm{E}(S_{it} \mid \tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it})$. We have*

$$F_\eta(\eta \mid \tilde{X}_{1it}, Z_{it}) = 1 - \pi(\tilde{X}_{1it}, Z_{it}, \ln P - \tilde{\beta}'\tilde{X}_{1it} - \eta),$$
$$F_W(w \mid X_{it}, Z_{it}) = 1 - \pi(\tilde{X}_{1it}, Z_{it}, \ln P - \ln w + \gamma'_c X_{2it,c}),$$

and $F_W(w \mid X_{it}, Z_{it}) = F_\eta(\ln w - \tilde{\beta}'\tilde{X}_{1it} - \gamma'_c X_{2it,c} \mid \tilde{X}_{1it}, Z_{it})$.

Proposition 3 not only provides closed form formulas of $F_W(w \mid X_{it}, Z_{it})$ in terms of CCP $\pi(\tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it})$, which can easily be estimated by a nonparametric regression of S_{it} on $\tilde{X}_{1it}, Z_{it}$ and $\tilde{Q}_{it}$, but also an approach to testing our model specifications. This result implies that the CCP $\pi(\tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it})$ as non-decreasing in explanable usage $\tilde{Q}_{it}$. To see this, we know that as a CDF, $F_W(w \mid X_{it}, Z_i)$ must be non-decreasing in w. By the formula of $F_W(w \mid X_{it}, Z_i)$ in terms of the CCP, we conclude that π is non-decreasing in usage for any given $(\tilde{X}_{1it}, Z_{it})$. Because we can always estimate the CCP function directly from the usage and subscription data, this property can be used in *model diagnostics*, to check our model specification and consistency with the assumptions. If we find that $\pi(\tilde{X}_{1it}, Z_i, \tilde{Q}_{it})$ is decreasing in usage $\tilde{Q}_{it}$ for certain intervals of $\tilde{Q}_{it}$ given $(\tilde{X}_{1it}, Z_{it})$, that would suggest that the model specification might be questionable (e.g., we have not controlled for enough consumer characteristics X_{it} or unobservable factors play a big role).

The CDF $F_W(W_{it} \mid X_{it})$ can be used to identify the demand curve.[3] First, one can obtain the unconditional distribution $F_W(W_{it}) = \mathrm{E}(F_W(W_{it} \mid X_{it}))$. Suppose p^c is a counterfactual price, we know that

$$D(p^c) \equiv \Pr(S_{it} = 1 \mid P = p^c) = \Pr(W_{it} > p^c) = 1 - F_W(p^c), \tag{7}$$

which can be understood as the demand function because it is the percentage of the population who would subscribe at price p^c. Knowing the demand function $D(p^c)$, we can identify the price elasticity of demand, which is essential for pricing.[4] When X_{it} includes consumer characteristics like gender, we can infer the demand function from $F_W(W_{it} \mid X_{it})$ in different market segments. For example, letting X_{it} be a categorical variable indicating gender,

$$\Pr(S_{it} = 1 \mid P = p^c, X_{it} = \text{Male}) = \Pr(W_{it} > p^c \mid X_{it} = \text{Male}) = 1 - F_W(p^c \mid X_{it} = \text{Male})$$

is the proportion of subscription among male consumers when price is p^c.

We note that the conditional WTP is more general than just obtaining the demand curve or unconditional WTP.[5] Next, we need to clarify what we *cannot* identify from data, and how that impacts the identification of WTP.

Proposition 4 (Unobservables and WTP). *The following holds regarding the identification of primitives:*

(i) The distribution of U_{it} and V_i cannot be separately identified, but rather the distribution of the sum $\eta_{it} = U_{it} + V_i$ is identified.

[3]When X_{it} are exogenous, Z_{it} vanishes from the identified $F_W(W_{it} \mid X_{it}, Z_{it})$. Otherwise, we have $F_W(W_{it} \mid X_{it}) = \mathrm{E}(F_W(W_{it} \mid X_{it}, Z_{it}) \mid X_{it})$, which is also identified.

[4]The price elasticity equals $D'(p^c)p^c/D(p^c)$, where $D'(p^c)$ is the first order derivative of $D(p^c)$.

[5]Our identification and estimation allows us to obtain the conditional WTP $F_W(W_{it} \mid X_{it})$. We might think of conditioning variables like demographics (age, gender, location etc.) or product characteristics (e.g. quality, features). It is important to note that our approach is not equivalent to the estimation of unconditional WTP $F_W(W_{it})$ or the demand curve $D(p)$, but rather more general.

There are two implications here. First, note that the conditional WTP is a more fundamental economic construct from which the unconditional WTP and demand curve can be derived by integrating it across the conditioning variable, whereas it is not possible to obtain the conditional WTP from either the unconditional WTP or demand curve. Second, the conditional WTP allows more flexibility and permits counterfactuals that would not be possible using the unconditional WTP or demand curve, $F_W(w)$ or $D(p)$. For example consider $F_W(W_{it} \mid EduAccount_{it})$, which provides the WTP distribution for educational and non-educational accounts. Thus, it would be possible to design a pricing plan with third-degree price discrimination, e.g. with educational discounts to the monthly music subscription, which would not be possible with just the demand curve. Similarly, if time-varying product characteristics are part of the conditioning variables, the conditional WTP would permit us to re-optimize pricing when product features are better.

(ii) *When X_{1it} and X_{2it} overlap, we can only identify the sum $(\beta_a + \gamma_a)$, and we do* not *separately identify β_a and γ_a.*

(iii) *The identification of WTP holds despite (i) and (ii) above.*

We find that the distribution of unobservables that affect usage, V_i, and WTP per unit usage, U_{it}, cannot be separately identified. From the usage panel data, we cannot identify unobserved heterogeneity in usage V_i due to the well known incidental parameters issue. Moving to the subscription choice, we then cannot tell whether a person did not subscribe because her unobserved usage heterogeneity V_i is small or because her unobserved WTP for one unit usage U_{it} is small.

Remark 1 (Can WTP per unit usage α_{it} be a function of expected usage Q^*_{it}?). It is intuitive to think WTP for subscription W_{it} as a function of expected usage Q^*_{it}. Then a natural question is that in our framework, is it possible to express the WTP per unit usage $\alpha_{it} = W_{it}/Q^*_{it}$ as a function of Q^*_{it}? The answer may appear to be yes but is actually no due to the required support conditions. Recall that in our framework, $\ln Q^*_{it} = \gamma' X_{2it} + V_i$ and $\ln \alpha_{it} = \beta' X_{1it} + U_{it}$. Making α_{it} a function of usage Q^*_{it} is essentially to make α_{it} a function of X_{2it} and V_i. Because we allow for correlation between U_{it} and V_i, one can just let U_{it} be equal to V_i. One may think let X_{1it} be equal to X_{2it}, so that α_{it} becomes a function of (X_{2it}, V_i), and hence a function of usage. However, this will violate the support condition in Assumption 2, which implies that we do need an excluded variable in X_{2it} that does not appear in X_{1it}. It may appear to be "yes" because WTP per unit usage α_{it} can be a function of many observable and unobservable factors affecting usage. However, it is "no" because of the "excluded variable" requirement. Our model covers at least two distinct dimensions (usage and WTP per unit usage) in analyzing WTP variation in the population.

Remark 2 (Difference from Heckman's Sample Selection Model). Our requirement for excluded variables are also needed in traditional sample selection model, e.g. Heckman's probit sample selection model (Heckman, 1979). Therefore, it is instrutive to clarify the difference between our framework and the sample selection model. In the sample selection model, there are two equations—selection equation (e.g. labor force participation decision) and outcome equation (e.g. earnings equation). In selection model, excluded variables that enter the outcome equation but not selection equation are required to implement reliable estimation. Readers may view our usage equation here as the outcome equation, and the subscription choice as the selection equation. Then, our excluded variables are indeed invalid in the context of traditional sample selection model, because the excluded variable herein indeed enters selection (i.e. subscription) equation—it just does not affect subscription apart from affecting usage (i.e. it affect WTP per unit usage). However, our estimation procedure is very different from the estimation of sample selection model. The estimation of a selection model has two steps. First, the selection equation is estimated. Second, using the estimation results from selection stage, the bias in estimating the outcome equation is corrected. In our approach, we need to do the reverse. First, we estimate usage equation, where we correct the sample selection using the feature of panel data. Then we use the estimated usage as a special regressor in estimating subscription decisions.

3.3 Estimation

We now detail how to carry out the estimation, following the previous arguments. The estimation recipe is detailed in Table 3 below. In Step 0, we collect the panel data for estimating WTP without price variation. In Step 1, we obtain the parameters γ governing usage, and obtain the conditional density of usage in Step 2. We then obtain the parameters β of the WTP or demand shifter using an IV regression in Step 3. In steps 4 and 5, we obtain the conditional CDF of WTP, from which we can also derive elasticities and the demand function.

Table 3: Estimation Recipe

Step	**Description**	**Details**
0	Data Requirements	Obtain panel data of n consumers over T billing periods. Data include price P, subscription choices S_{it}, usage Q_{it} when they subscribe, observable factors that impact WTP per unit usage (X_{1it}) and usage (X_{2it}). When X_{1it} is endogenous, we also need IV Z_{it} for X_{1it}.When X_{1it} is not endogenous, $Z_{it} = X_{1it}$.
1	Fixed Effect Estimation on Usage	Using the data of subscribers, estimate the following panel data model $\ln Q_{it} = \gamma_a' X_{2i,a} + \gamma_b' X_{2i,b} + \gamma_c' X_{2it,c} + V_i + \varepsilon_{it}$ with fixed effect estimator. Let $\hat{\gamma}_c$ be the estimator of γ_c, and let $\hat{Q}_{it} = \hat{\gamma}_c' X_{2it,c}$ for all i (subscribers and non-subscribers) and t.
2	Nonparametric Estimation of $f_{\tilde{Q}}(\tilde{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})$	Obtain a nonparametric estimator $\hat{f}_{\tilde{Q}}(\cdot \mid \tilde{X}_{1it}, Z_{it})$ of the conditional PDF $f_{\tilde{Q}}(\tilde{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})$ from the sample $(\hat{Q}_{it}, \tilde{X}_{1it}, Z_{it})$.
3	IV Regression to Estimate $\tilde{\beta}$	Estimate $\tilde{\beta}$ by the 2SLS regression of $\hat{Y}_{it}$ on $\tilde{X}_{1it}$ using IV Z_{it}, where $\hat{Y}_{it} = (S_{it} - \mathbb{I}(\hat{Q}_{it} - \ln P > 0))/\hat{f}_{\tilde{Q}}(\hat{Q}_{it} \mid \tilde{X}_{1it}, Z_{it})$.
4	CCP Estimation of π	Estimate the CCP function $\pi(\tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it}) = \mathrm{E}(S_{it} \mid \tilde{X}_{1it}, Z_{it}, \tilde{Q}_{it})$ from the sample $(S_{it}, \tilde{X}_{1it}, Z_{it}, \hat{Q}_{it})$.
5	Estimate CDF of WTP	Estimate the conditional CDF of WTP F_W and of unexplained heterogeneity in WTP F_η using the formulas in Proposition 3.

Remark 3 (Desirable Features). Our approach presented above has a number of desirable features. First, we do not assume that for an individual consumer, WTP for a service is constant over time. There might be a number of reasons why the constant WTP may not be true, which could depend on the product variety available (e.g. number of songs on streaming service) or change in unobservable quality (e.g. with a gym renovation). In addition, persistent unobservable heterogeneity is well-recognized to be important in the marketing and economics literature. With individual level data, these are typically included. In our framework, we examine both the heterogeneity in usage and WTP per unit usage. These two sources of heterogeneity are distinct but related. In the gym example, the model with these two dimensional heterogeneity can accommodate two types

of gym members: subscribers who expect to use gym often like college students, and subscribers who are willing to pay more for one gym visit though they may not use gym often like professional lawyers.

Second, it is important to allow for correlation between usage and WTP per unit usage, which leads to correlation between usage and WTP for subscription. It's important *not* to assume that usage differences across consumers reflect WTP differences. Whereas we might intuitively expect usage to be positively correlated with WTP, that argument only holds *within a consumer*, but not across consumers. For example, a college student might spend a lot of time listening to a music streaming service, or visiting a gym, whereas a professional lawyer might spend little time on either. However, it might well be the case that the WTP for the professional is much higher than the WTP for the college student, reflecting a negative correlation between usage and WTP. Thus, we need a flexible model to accommodate arbitrary correlations between usage and WTP per unit usage.

Third, there is the question of whether our goal is to obtain the distribution of WTP of subscribers or more generally of the population. In commonly used approaches, the random coefficients are assumed to be drawn from some known distribution, and then the researchers are thus able to recover the responses from potential consumers who are not observed at all in the data. We specify consumer selection into purchases using a selection model, which also allows for unobservables to impact whether a consumer ever purchases (and whose usage is thus available in our data). In counterfactuals, our method makes explicit the assumptions regarding the population-level observables that are required to extrapolate from the observed set of consumers to the population of potential consumers. If the researcher is only interested in examining the impact on customers who are present in the data, then such additional data is not required.

Fourth, our framework does not require us to assume that unobserved heterogeneity (in either usage or WTP per unit usage) has a known distribution, such as the standard normal or type I extreme value distribution, which are typically made in the literature. This approach makes it less susceptible to specification error, when purchase and usage decisions are heavy-tailed or multi-peaked, for example.

3.4 Unobservable Product Quality

There could be two sources of the unobserved heterogeneity U_{it} in the valuation about the subscribed service. One is time invariant consumer specific unobserved heterogeneity, and the other is the unobserved product characteristic that is common to all consumers.[6] We let ζ_i and ξ_t to denote the former and the latter, respectively, and let $U_{it} = \zeta_i + \xi_t$. Then $\eta_{it} \equiv U_{it} + V_i = \omega_i + \xi_t$, where $\omega_i \equiv \zeta_i + V_i$. Below, we are going to show that with panel data we indeed can separately identify the distributions of ω_i and ξ_t under certain conditions. To simplify the exposition, we omit X_{1it} and assume $\tilde{Q}_{it} \perp\!\!\!\perp \eta_{it}$ below.

[6] Unobserved product characteristic of course can also affect usage, but this can be easily addressed by including period dummy variables in our usage panel data model.

The idea is to note that ω_i is time invariant in the above specification. Below, we first identify the values of the difference $\xi_{t+1} - \xi_t$ by the time variation (eq. (8) below), which is possible when the number of consumers n is large. Thus, by assuming an initial value ξ_0, we can identify each value of ξ_t, hence the distribution of ξ_t. Then the distribution of ω_i is easily obtained from the distribution of $\eta_{it} - \xi_t$ since we have identified the distribution of η_{it} using Proposition 3 and the value of ξ_t is known.

As an alternative of assuming an inital value ξ_0, one can assume that ξ_t is serially independent and identically distributed, we then can identify its distribution from the difference $\xi_{t+1} - \xi_t$ by using the constrained deconvolution (Belomestnyi, 2002). The constraint is that $F(\xi_t) = F(\xi_{t+1})$. After identifying the distribution $F(\xi_t)$, we can identify the distribution of ω_i by deconvolution of the previously identified distribution of $\eta_{it} = \omega_i + \xi_t$ from assuming that $\xi_t \perp\!\!\!\perp \omega_i$.

Proposition 5 (Can Identify: Unobservable Product Quality)**.** *With fixed distribution of unobservable product quality over time, both the distribution of unobserved quality and the individual fixed effects can be identified. The differences in unobservable product quality is given by:*

$$\xi_{t+1} - \xi_t = \mathrm{E}_{t+1}(\tilde{Y}_{i,t+1}) - \mathrm{E}_t(\tilde{Y}_{it}). \tag{8}$$

where $Y_{it} = (S_{it} - \mathbb{I}(\tilde{Q}_{it} - \ln P > 0))/f_{\tilde{Q},t}(\tilde{Q}_{it})$*. Here* $f_{\tilde{Q},t}(\tilde{Q}_{it})$ *is the density function of* $\tilde{Q}_{it}$ *in period* t*.*

Estimation is straightforward since the right-hand-side (RHS) of eq. (8) is estimable by sample averages.

4 Where is Price Variation Useful?

Our previous analysis has focused on a case where there was no price variation, which was the primary setting of interest. While our prior results have identified how the combination of subscription choice data and usage data can help identify the WTP distribution, here we demonstrate that having such data is not equivalent to settings that feature price variation.

We had previously focused on a case where WTP was of a multiplicative form: $W_{it} = \alpha_{it} Q^*_{it}$. One natural question that arises is whether we can recover more general functional forms for which WTP depends on expected quantity. The answer to this question depends on whether or not we have price variation at all in data.

Proposition 6 tells us that without any price variation, observing subscription/churn choices and usage tracking data cannot permit us to identify the functional form of WTP W_{it} as a function of usage Q^*_{it}. This negative result provides a useful "boundary condition" for the previous results. However, for any known functional form, we can recover the parameters that characterize the distribution. This could be important since without knowing the functional form, the distribution of the WTP cannot be identified.

Proposition 6 (Cannot Identify Functional Form of WTP Without Price Variation)**.** *Suppose the WTP is specified as* $W_{it} = g(\ln \alpha_{it} + \ln Q^*_{it})$*, where* g *is a strictly increasing function.*[7] *In the absence of price variation, the distribution of WTP cannot be identified without specifying the functional form* g*.*

This proposition is hardly surprising after noting that W_{it}, hence observed subscription choice, depends on two unknown functions: g and the demand shifter α_{it}. When there are two unknowns, there are multiple ways to assign the pair (g, α_{it}) so that the resulted W_{it}, hence observed subscription choice, stays unchanged.

4.1 Benefits of Price Variation Across Markets

We have seen that certain tight specification, like multiplicative form is necessary to identify the distribution of the WTP. We are going to argue that the price variation can help to relax this specification, and the extent of the relaxation depends on the observed variation of price.

When we observe data from multiple markets, it is possible that the same subscription service has different prices. For example, the PlayStation Plus (Sony's prime service for PlayStation players) costs \$59.99 annually plus tax in the US and £49.99 annually in the UK.

Suppose in the sample, there are M markets across which the price varies. Let P_m be the price in market m, and the other notation including $W_{it,m}$, $X_{1it,m}$, $X_{2it,m}$, and $\tilde{Q}_{it,m}$ are defined similarly for each market m. For the simplicity of exposition, suppose $X_{2it,m}$ is all time varying, then $\tilde{Q}_{it,m} = \gamma' X_{2it,m}$. Assume that the WTP for the service in market $m = 1, \ldots, M$ equals,

$$W_{it,m} = g(\beta' X_{1it,m} + \tilde{Q}_{it,m} + U_{it,m} + V_{i,m}; \delta) = g(\beta' X_{1it,m} + \tilde{Q}_{it,m} + \eta_{it,m}; \delta).$$

Let $g(\cdot; \delta)$ be a strictly increasing function known up to a finitely dimensional vector of parameters δ. We have a concrete example in the next section. Assume $\tilde{Q}_{it,m}$ has been identified. The subscription decision $S_{it,m}$ will depend on the local price P_m in the market m where consumer i resides. We then have

$$S_{it,m} = \mathbb{I}(\beta' X_{1it,m} + \tilde{Q}_{it,m} + \eta_{it,m} > g^{-1}(P_m; \delta)). \tag{9}$$

For simplicity, assume that $X_{1it,m}$ is exogenous, so Z_{it} can be ignored below. By the same arguments in the previous sections, we have

$$\beta = \mathrm{E}(X_{1it,m} X'_{1it,m})^{-1} \mathrm{E}(X_{1it,m} Y_{it,m}),$$

where

$$Y_{it,m} = \frac{S_{it,m} - \mathbb{I}(\tilde{Q}_{it,m} - g^{-1}(P_m; \delta) > 0)}{f_{\tilde{Q}_m}(\tilde{Q}_{it,m} \mid X_{1it,m})}.$$

[7]The multiplicative form $W_{it} = \alpha_{it} Q^*_{it} = \exp(\ln \alpha_{it} + \ln Q^*_{it})$ assumes that g is an exponential function.

Suppose $X_{1it,m}$ has the same distribution across different markets, hence $\mathrm{E}(X_{1it,m}X'_{1it,m})$ is identical across m. We then have a set of moment equations of δ,

$$\mathrm{E}(X_{1it,j}Y_{2it,j}) - \mathrm{E}(X_{1it,k}Y_{2it,k}) = 0, \qquad 1 \le j < k \le M, \tag{10}$$

for any pair of two markets (j,k). The number of moment equations depends on the dimension of X_{1it} and the number of markets. When there are enough number of markets with different prices, we can identify the parameters δ of $g(\cdot;\delta)$. Once g is identified, the identification of the distribution of W_{it} follows from the earlier results.

4.2 How to Reconcile Zero Usage among Subscribers?

Consumers may just purchase for the option to use the service, or have access to the service although they may not the service. Thus, we might observe cases in the data where we see *zero usage* of the service. Note that we need to allow for the possibility that one's WTP for subscription can still be positive and perhaps higher than charged price even if they don't use the subscribed product.

We can consider the specification $W_{it} = \mu + \alpha_{it}Q^*_{it}$. Note that even if $Q^*_{it} = 0$, a consumer still have positive valuation μ. As one application of the above arguments about the role of price variation, we show how to identify μ by price variation within the same market. The distinct prices for the *same subscription plan* can come from exogenous price changes of the subscription plan across time. For example, the PlayStation Plus annual subscription in the UK was priced at £39.99 before September 2017, and it costs £49.99 now.

Suppose there are $P_1, \ldots, P_M$ prices of the same subscription in a market. Consumers facing price P_m make the subscription decision based on the rule $S_{itm} = \mathbb{I}(W_{it,m} > P_m)$ that is equivalent to $S_{itm} = \mathbb{I}(\ln(\alpha_{it}Q^*_{it}) > \ln(P_m - \mu))$. Using $\ln(\alpha_{it}Q^*_{it}) = \beta' X_{1it,m} + \tilde{Q}_{it,m} + \eta_{it,m}$, we can write

$$S_{it,m} = \mathbb{I}(\beta' X_{1it,m} + \tilde{Q}_{it,m} + \eta_{it,m} > \ln(P_m - \mu)).$$

Here the valuation of the service without using it (μ) is unknown, but invariant of the prices. This specification is a special case of eq. (9) by letting $g(c;\delta) = e^c - \delta$ and $\delta = -\mu$. We then can use the moment eq. (10) to identify μ. To be concrete, suppose $X_{1it,m} = 1$ only. Equation (10) reads

$$\mathrm{E}\left(\frac{S_{it,j}}{f_{\tilde{Q}_j}(\tilde{Q}_{it,j})} - \frac{S_{it,k}}{f_{\tilde{Q}_k}(\tilde{Q}_{it,k})}\right) = \mathrm{E}\left(\frac{\mathbb{I}(\tilde{Q}_{it,j} > \ln(P_j - \mu))}{f_{\tilde{Q}_j}(\tilde{Q}_{it,j})} - \frac{\mathbb{I}(\tilde{Q}_{it,k} > \ln(P_k - \mu))}{f_{\tilde{Q}_k}(\tilde{Q}_{it,k})}\right), \tag{11}$$

for any $1 \le j < k \le M$.

To see more intuition, it can be shown that

$$\text{RHS of eq. (11)} = \ln\left(\frac{P_k - \mu}{P_j - \mu}\right),$$

and

$$\text{LHS of eq. (11)} = \int_{\mathcal{Q}} \int_{\eta} (S_{it,j} - S_{it,k}) \, \mathrm{d} F_{\eta}(\eta) \, \mathrm{d} q.$$

The left-hand-side (LHS) can be interpreted as the difference between the percentage of subscription in two markets as if the log of usage ($\tilde{Q}_{it,m}$) was uniformly distributed. The moment equation links the change of subscription percentage to the change of price across markets.

5 Numerical Simulation

We show numerical performance of our proposed method. The model specification follows the discussion above, with $\ln \alpha_{it} = \beta_1 + \beta_2 X_{1it} + U_{it}$. Data were generated from the following:

$$\begin{aligned} S_{it} &= \mathbb{I}(\beta_1 + \beta_2 X_{1it} + \gamma_1 + \gamma_2 X_{2it} - \ln P + V_i + U_{it} > 0) \\ &= \mathbb{I}(0.5 + 1 \cdot X_{1it} + 0.5 + 1 \cdot X_{2it} - \ln P + V_i + U_{it} > 0), \\ \ln Q_{it} &= \gamma_1 + \gamma_2 X_{2it} + V_i + \varepsilon_{it}, \end{aligned}$$

where

$$\begin{aligned} V_i &= \nu_{1i}, & U_{it} &= \nu_{1i} + \nu_{2it}, & X_{1it} &= \nu_{1i} + \nu_{3it}, \\ Z_{it} &= (1, \nu_{3it})', & X_{2it} &= \nu_{4it}, & \varepsilon_{it} &= \nu_{5it}. \end{aligned}$$

Here ν's are mutually independent normally distributed random variables. Observe that ν_{1i} is included in both U_{it} and V_i, so they are correlated. Similarly, X_{1it} is endogenous (correlated with both V_i and U_{it}), Z_{it} is IV for X_{1it}, and X_{2it} is exogenous whose support conditional on (X_{1it}, Z_{it}) is real line. We specify $\nu_{1i} \sim \mathcal{N}(0, 0.5^2)$, $\nu_{2it} \sim \mathcal{N}(0, 0.5^2)$, $\nu_{3it} \sim \mathcal{N}(1, 1)$, $\nu_{4it} \sim \mathcal{N}(1, \sigma_{x_2}^2)$, $\nu_{5it} \sim \mathcal{N}(0, 1)$. By this construction, $\eta_{it} = U_{it} + V_i$ and $\eta_{it} \sim \mathcal{N}(0, 1.25)$. It is easy to show that $\eta_{it} \mid X_{1it} \sim \mathcal{N}(0.4(x_1 - 1), 1.05)$. Hence the $W_{it} \mid (X_{1it} = x_1, X_{2it} = x_2)$ follows a log normal distribution, and the mean and variance of its logarithm are $1.4x_1 + x_2 + 0.6$ and 1.05, respectively. We let $\ln P = 2$.

One important condition is that the support of $\tilde{Q}_{it,m} \mid X_{1it}$ covers the support of $\ln P - \beta' X_{1it} - \eta_{it} \mid X_{1it}$. In this numerical example, this condition requires that the support of $X_{2it} \sim \mathcal{N}(1, \sigma_{x_2}^2)$ covers the support of $1 - X_{1it} - \eta_{it} \mid X_{1it} \sim \mathcal{N}(1.4(1 - X_{1it}), 1.05)$. Whether or not this support condition holds depends on the value of X_{1it} and σ_{x_2}. Figure 2 displays the 95% confidence interval as well as the density function of X_{2it} and $1 - X_{1it} - \eta_{it} \mid X_{1it}$ when $\sigma_{x_2} = 0.5$, 1 and 1.5, and X_{1it} equals to 10%, 25%, 50%, 75% and 90% percentile of its marginal distribution. We observe that when σ_{x_2} is too small ($\sigma_{x_2} = 0.5$), the support condition hardly holds. When we increase σ_{x_2}, it is easier to satisfy the support condition.

All reported results were based on 250 replications. Table 4 shows the estimation of β_2, γ_2 and $\beta_1 + \gamma_1$. Note that we cannot separately identify β_1 and γ_1. It is interesting to observe that

Figure 2: Illustrate Support Condition

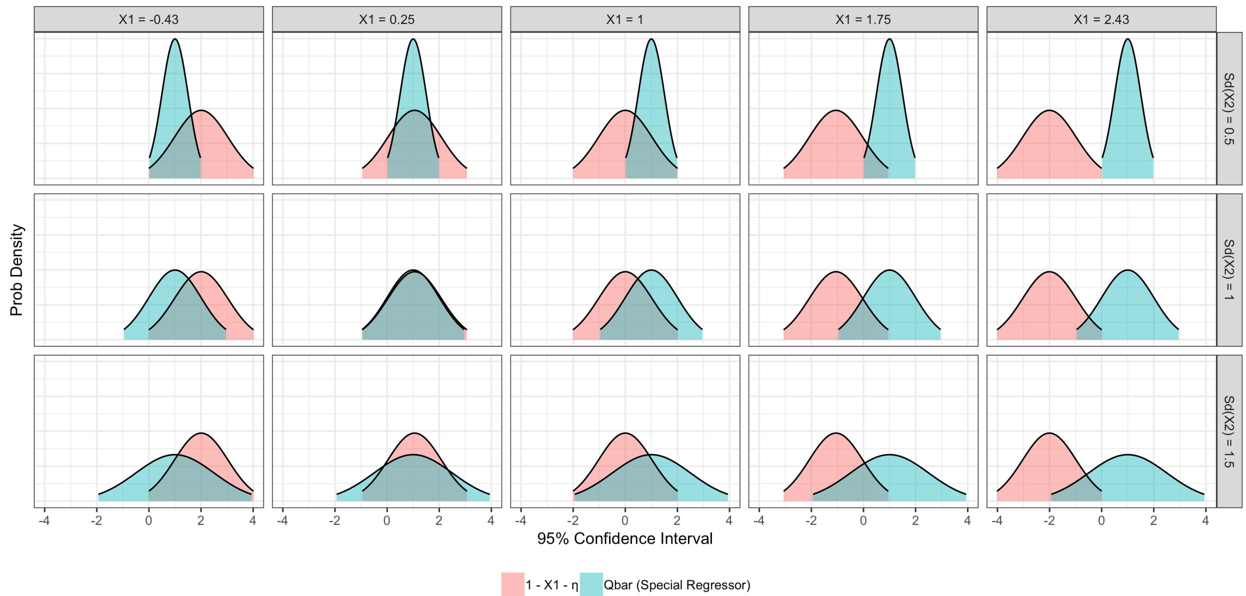


Table 4: Parameter Estimates: $T = 2$

	$n = 1000$			$n = 2000$		
	$\beta_2 = 1$	$\gamma_2 = 1$	$\beta_1{+}\gamma_1 = 1$	$\beta_2 = 1$	$\gamma_2 = 1$	$\beta_1{+}\gamma_1 = 1$
$\sigma_{x_2} = 0.5$	0.551	1.004	1.005	0.529	0.992	1.029
	(0.159)	(0.086)	(0.181)	(0.124)	(0.065)	(0.148)
$\sigma_{x_2} = 1.0$	0.824	0.999	1.074	0.846	0.999	1.061
	(0.181)	(0.048)	(0.145)	(0.160)	(0.034)	(0.180)
$\sigma_{x_2} = 1.5$	0.962	1.002	1.048	0.962	1.000	1.066
	(0.208)	(0.036)	(0.253)	(0.162)	(0.023)	(0.100)

Note: Results are based on 250 replications. Standard deviation is in parenthesis.

when σ_{x_2} is small making the support condition harder to withstand, there is significant bias in estimating β_2. It is also interesting to observe that when $\sigma_{x_2} = 1.5$ and the support condition does not strictly hold (see Figure 2), the bias in estimating β has became negligible. The effects of support conditions are further illustrated by the CDF estimation in Figure 3 and the estimation of $\mathrm{E}(\ln W_{it} \,|\, X_{1it}, X_{2it})$ in Figure 4. Comparing the bias of estimating $\mathrm{E}(\ln W_{it} \,|\, X_{1it}, X_{2it})$ in Figure 4 with the support condition Figure 2, one can see that the bias becomes larger when it becomes harder to satisfy the support condition.

6 Empirical Application: Music Streaming Service Subscription

In this empirical application, we focus on the market of online music streaming service in Southeast Asia during the period January 2015–February 2017. We represent the prices in scaled \$ terms

Figure 3: Estimation of the CDF of $\eta \mid X_{1it} = 1$ and $W_{it} \mid (X_{1it} = 1, X_{2it} = 1)$

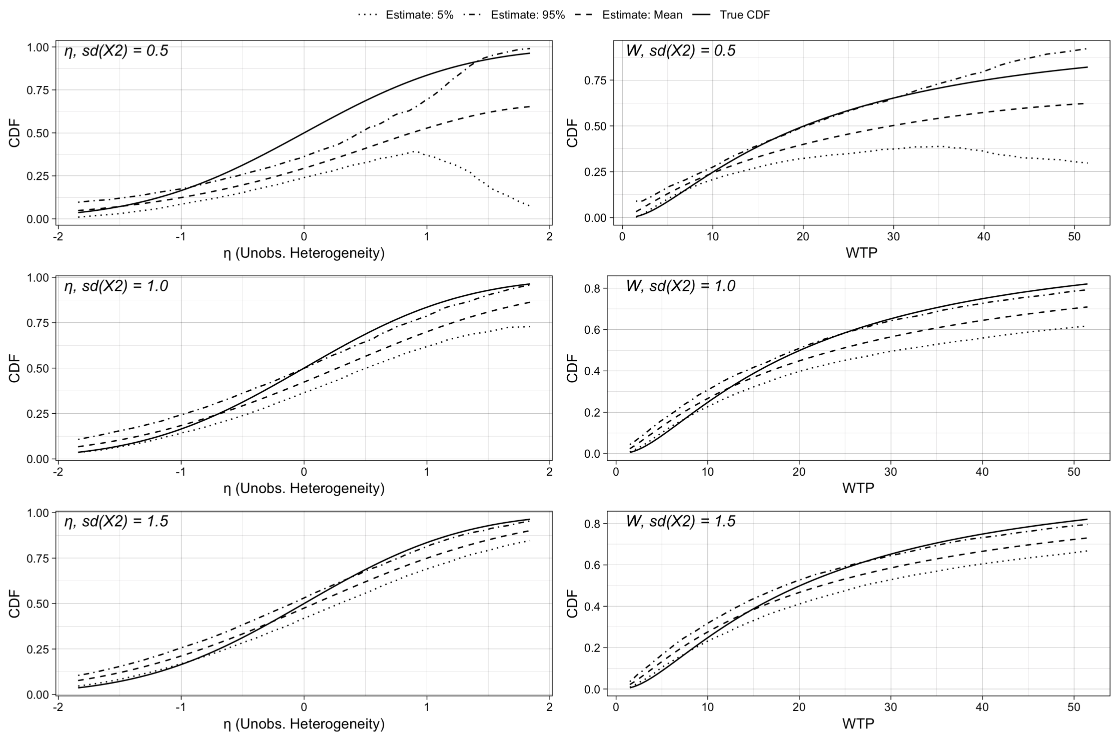
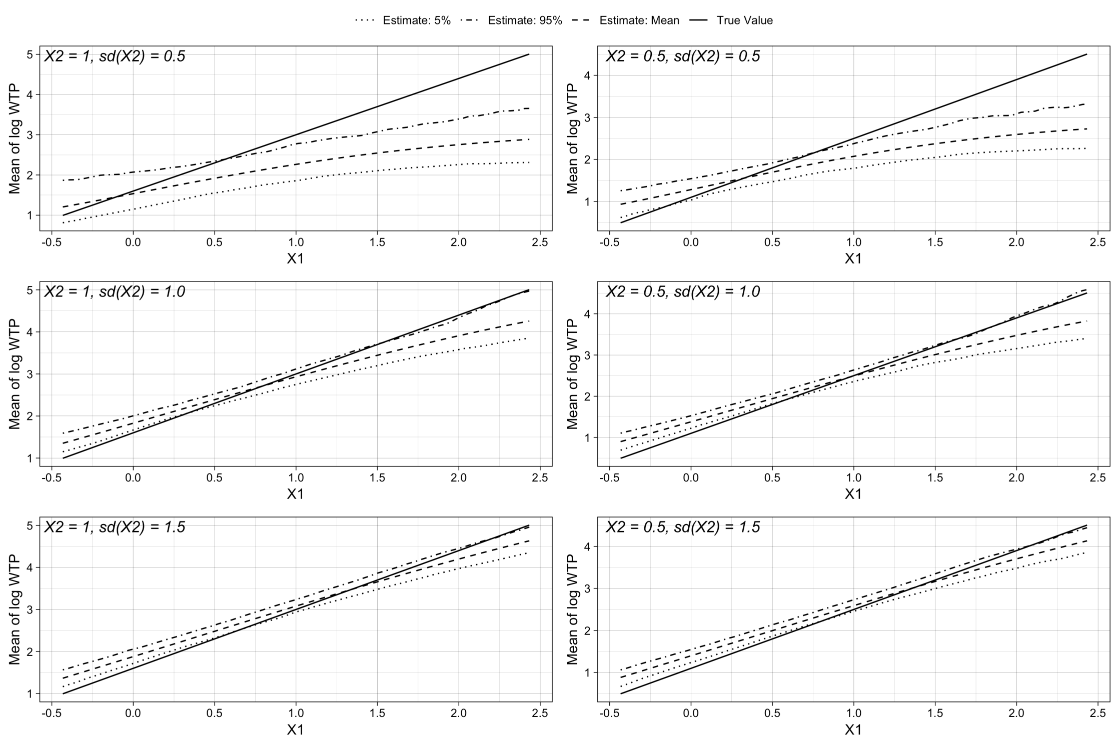


Figure 4: Estimation of $\mathrm{E}(\ln W_{it} \mid X_{1it}, X_{2it})$



for exposition and to avoid attribution to the firm that provided the data. The usage data are unscaled.

We examine an empirical setting in which we study the churn decision of a customer, and we use our method to obtain estimates of the elasticities of the WTP to age and gender, the mean and CDF of the log WTP for the monthly streaming service.

Data

The data were provided by a music streaming service company, which targets the Southeast Asian market. Its service has 80 percent market share. Though the company sells subscription plans of varying lengths (e.g., Monthly, 180 days, 365 days), most users (93.7 percent in our sample) choose monthly plan. Registered users can also listen to music free for up to 1 hour each day with various restrictions, however only 3.56 percent of the users in our sample have ever used this free service. We will focus on the subscription choice of the monthly plan, which has three different prices in our sample. The first price is its listed price \$149 and is the baseline price. Special prices of \$129 and \$99 are only available to the customers of certain VISA credit cards and cellular carriers. In our sample, the percentages of the three prices \$149, \$129, and \$99 are 56.8, 17.7 and 23.9, respectively. Note that these prices are for the *same* monthly music streaming service.

The raw data include more than 1 million registered users. We observe the monthly usage (the number of seconds each user listened to music with the service) of its subscribers. We also observe each user's payment transaction history during that period. Thus, we know which users have churned, or left the service.[8] In terms of demographics, we can only observe age and gender.[9] We randomly sampled 8698 users for the analysis below.

Table 5 shows the basic summary statistics of the market by summarizing customers' usage and characteristics. We can see that there is no significant difference in term of age and gender between the customers who churned and who did not. Examining monthly usage, we find that customers who churned used the service less than those who did not. Figure 5 further illustrates this observation by plotting the kernel density functions of the average and the log of the average monthly listening hours.

Model and Estimation

For $p \in \{\$99, \$129, \$149\}$, let (i, p) denote a consumer who pays p for her monthly plan. For each customer i, let $S_{i,p}$ denote her churn decision, and $S_{i,p} = 0$ if she churns and $S_{i,p} = 1$ otherwise.

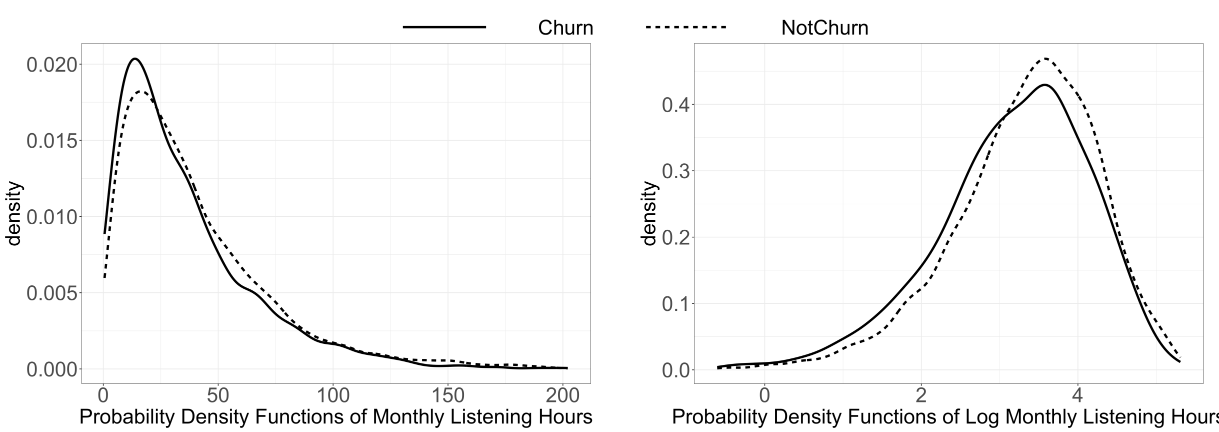
[8] We define a customer as churning if there is no new valid service subscription within 30 days after the his or her current membership expires. According to this definition, only 2.5 percent of customers who have churned would come back to use the service some time later.

[9] The age and gender of some users are missing. Comparing the kernel densities of the usage by those with and without age and sex information, we did not find significant difference, hence we removed the user without valid age and gender observations.

Table 5: Descriptive Statistics of Music Streaming Service Sample

Panel A: Users who churned ($n = 1211$)						
	Mean and std. dev.	First Quartile	Median	Third Quartile	Min	Max
Average monthly listening hours (lifetime in sample)	35.22 (29.80)	13.53	26.44	48.36	0.59	200.78
Tenure (weeks)	139.82 (93.06)	55.93	130.57	210.93	0.14	349.57
Age	29.53 (8.25)	23.00	28.00	34.00	17.00	59.00
Is Female	0.49					
Panel B: Users who did not churn ($n = 7487$)						
	Mean and std. dev.	First Quartile	Median	Third Quartile	Min	Max
Average monthly listening hours (lifetime in sample)	39.95 (32.55)	16.53	31.21	54.27	0.54	202.12
Tenure (weeks)	164.40 (99.89)	68.43	162.86	252.43	0.14	376.86
Age	29.92 (8.14)	24.00	28.00	34.00	17.00	59.00
Is Female	0.45					

Figure 5: Kernel Density Functions of (Log) Average Monthly Listening Hours



Her subscription choice is based on:

$$S_{i,p} = \mathbb{I}(\ln W_{i,p} > \ln(p + \mu)),$$

where $W_{i,p}$ is her WTP for the monthly music streaming service, and μ is the money-metric expected utility of the outside option. We specify the WTP $W_{i,p}$ as:

$$\ln W_{i,p} = \ln \alpha_{i,p} + \ln Q^*_{i,p} = \beta_1 + \beta_2 Age_{i,p} + \beta_3 Female_{i,p} + \ln Q^*_{i,p} + U_{i,p},$$

where $Female_{i,p}$ is a female dummy variable, and $Q^*_{i,p}$ is the expected monthly usage. For users who churn, we let $Q^*_{i,p}$ be the total listening hours during the last month before churn. For users who did not churn, we let $Q^*_{i,p}$ be the average monthly listening hours in their subscription period.

Our framework still applies in the case where expected usage Q^*_{it} is observed by letting $X_{2it} = \ln Q^*_{it}$, $\gamma = 1$, and $V_i = 0$, hence $\tilde{Q}_{it,m} = \ln Q^*_{it}$. Here, we use this simplified specification. Let $\tilde{Q}_{it} = \ln Q^*_{it}$. Moreover, we assume $Age_{i,p}$ and $Female_{i,p}$ are exogenous, so that we can omit IV Z_{it}.

Our estimation follows the following steps. First, we use a nonparametric kernel density estimator to estimate the conditional PDF $f_{\ln Q^*}(\ln Q^*_{i,p} \,|\, Age_{i,p}, Female_{i,p})$ using the sample of consumers who pay price p. Denote $\hat{f}_{\ln Q^*}$ the estimator, and let $\hat{f}_{i,p} \equiv \hat{f}_{\ln Q^*}(\ln Q^*_{i,p} | Age_{i,p}, Female_{i,p})$. Second, because we have three prices in data, we use the moment eq. (11) to estimate the money-metric expected utility μ of the outside option. In particular, we find the minimizer of

$$\hat{\mu} = \arg\min_{\mu > 0} \hat{g}(\mu)' \hat{g}(\mu),$$

where $\hat{g}(\mu) = (\hat{g}_{129}(\mu), \hat{g}_{99}(\mu))'$, and for $p = 99, 129$ the moment function $\hat{g}_p(\mu)$ is defined as follows,

$$\hat{g}_p(\mu) = \left[\frac{1}{n_p} \sum_{(i,p)} \frac{S_{i,p}}{f_{i,p}} - \frac{1}{n_{149}} \sum_{(i,149)} \frac{S_{i,149}}{f_{i,149}}\right] - \left[\frac{1}{n_p} \sum_{(i,p)} \frac{\mathbb{I}(\ln Q^*_{i,p} - \ln(p+\mu) > 0)}{f_{i,p}} - \frac{1}{n_{149}} \sum_{(i,149)} \frac{\mathbb{I}(\ln Q^*_{i,149} - \ln(149+\mu) > 0)}{f_{i,149}}\right],$$

where n_p is the number of consumers who paid p for their monthly plan in the sample. The estimated money-metric expected utility of the outside option is $\hat{\mu} = 38.15$ with standard error 9.24. The standard error was obtained from bootstrap. Third, using the data of subscribers who paid \$149 for their monthly plan, we estimate $\beta = (\beta_1, \beta_2, \beta_3)'$ by linear regression of

$$Y_{i,149} = \frac{S_{i,149} - \mathbb{I}(\ln Q^*_{i,149} - \ln(149 + \hat{\mu}) > 0)}{\hat{f}_{i,149}},$$

on $Age_{i,149}$ and $Female_{i,149}$. Fourth, we estimate nonparametrically

$$\pi(Age_{i,149}, Female_{i,149}, \ln Q^*_{i,149}) \equiv \mathrm{E}(S_{i,149} \mid Age_{i,149}, Female_{i,149}, \ln Q^*_{i,149}).$$

Table 6: WTP and Usage Elasticities (Hundredths) Estimates

	WTP	Usage (All Users)	Usage (Churn)	Usage (Not Churn)
Age	3.98	−2.00	−2.18	−2.09
	(1.19)	(0.22)	(0.65)	(0.20)
Is female	−26.23	−3.13	9.49	−4.75
	(18.09)	(3.46)	(10.34)	(3.14)
n	4159	4159	913	3246

Then the CDF of the WTP is obtained from the formula in Proposition 3.

Results

Table 6 reports the estimated elasticities of the WTP and usage to age and gender using the following slight modification of Proposition 2,

$$\mathrm{E}(\ln W_{i,p} \mid Age_{i,p}, Female_{i,p}) = \beta_1 + \beta_2 Age_{i,p} + \beta_3 Female_{i,p} + \mathrm{E}(\ln Q^*_{i,p} \mid Age_{i,p}, Female_{i,p}).$$

We approximated $\mathrm{E}(\ln Q^*_{i,149} \mid Age_{i,149}, Female_{i,149})$ with a linear regression of $\ln Q^*_{i,149}$ on $Age_{i,149}$ and $Female_{i,149}$. It is interesting to see that even though age elasticity of the usage is negative, age elasticity of the WTP is significantly positive. This is likely due to the fact that age is positively correlated with earnings, and an older user might be willing to pay more for one unit usage (e.g. one hour of music). Their usage might be low due to opportunity costs of time. However, their WTP for listening to one hour of music is high enough so that the WTP for the monthly service is greater despite their lower usage relative to younger users. The effects of gender on the WTP is not significant (with p-value 0.13), though the sign of the elasticity may be due to different income of men and women in the sample.

Table 7 reports the estimated mean of log of WTP $\mathrm{E}(\ln W_{i,149} \mid Age_{i,149}, Female_{i,149})$ and mean of log monthly number of listening hours. We divided age into four categories, and estimated the mean for each category and gender. We observe that on average, one is willing to pay substantially more than the current listed price $\ln(149) \approx 5$ regardless of age and gender. This explains the high renewal rate (more than 93 percent in sample) in a given month.

We do not observe a gender difference in the WTP for age ranges before college. This is likely because these high school students have not started earning any income, so the gender earnings gap does not affect the WTP. For the age between 19 and 30, men's average WTP is higher than women's.

Figure 6 shows the estimated CDF of the log WTP of a male user with usage at 10 and 50 percentile of the population. Looking at the median, we note that WTP seems quite elastic to his usage.

We estimated the demand curve using the demand function eq. (7). Figure 7 shows the estimated

Table 7: Log of WTP and Usage Estimates

	Mean of Log WTP		Mean of Log Monthly Hours	
	Men	Women	Men	Women
Age $\leq$ 18 (before college)	6.2	6.2	ln(26.6 Hrs)	ln(25.8 Hrs)
Age between 19 and 22 (college)	6.6	6.5	ln(31.6 Hrs)	ln(27.6 Hrs)
Age beween 23 and 30	6.9	6.8	ln(29.1 Hrs)	ln(26.5 Hrs)
Age > 30	7.1	7.3	ln(19.4 Hrs)	ln(21.1 Hrs)

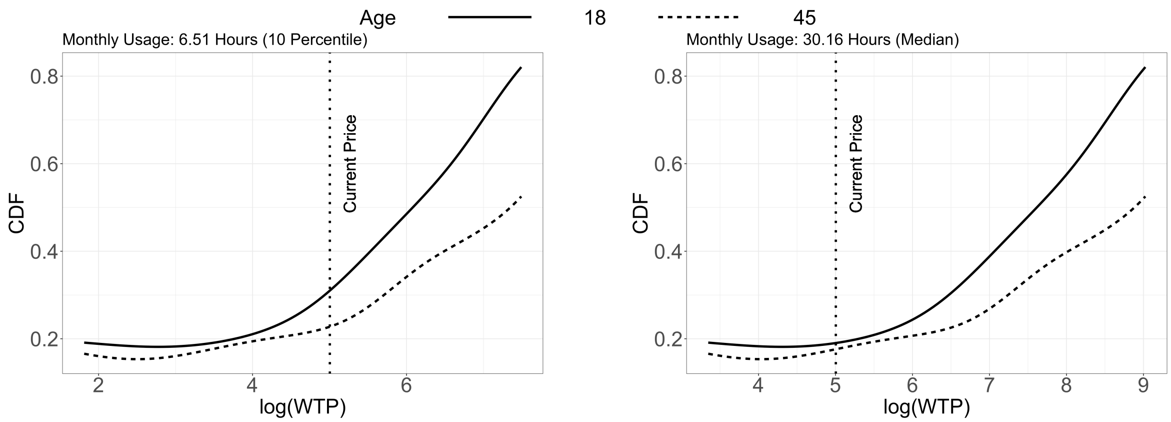


Figure 6: Estimated CDF of the log WTP of Men

demand curve. The y-axis is the counterfactual price, and the x-axis is the percentage of the current customers who would pay such counterfactual price and keep subscribing the monthly service. The estimated curve suggests that consumers are very inelastic at the current price. One caveat is that in estimating the demand curve, we rely on

$$\Pr(S^c_{i,149} = 1 \mid X_{1i}, Q^*_{i,149}, P^c) = \pi_1(X_{1i}, \ln(149 + \mu) - \ln(P^c + \mu) + \ln Q^*_{i,149}),$$

where the superscript c denotes a counterfactual object. Note that the probability $\Pr(S^c_{i,149} = 1 \mid X_{1i}, Q^*_{i,149}, P^c)$ is estimable only when $\ln(P + \mu) - \ln(P^c + \mu) + \ln Q^*_{i,149}$ is within the support of observed log of usage in data (which is generated under the current price). When the counterfactual price P^c differs greatly from the current price, such a support condition requirement may not hold, and thus the obtained demand $\Pr(S^c_{i,149} = 1 \mid P^c)$ may be unreliable. This implies that one may not be able to identify the optimal price when the usage variation is not big enough, but the firm can at least identify whether or not the current price is too high or low because the slope of the profit function at the current price is identified when the marginal cost is known.

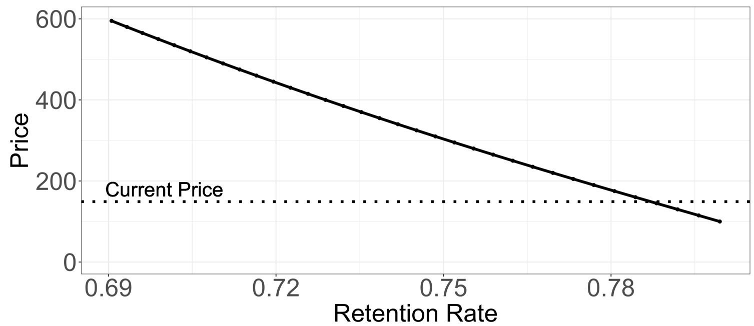


Figure 7: Estimated Demand Curve

7 Conclusion

Many subscription commerce markets charge the same price for every consumer and over time. Thus, price variation is very limited, if any. In such cases, classic results and arguments from the literature discuss how the identification of demand or WTP is not possible without price variation.

Our research suggests that if one is willing to impose certain functional form of the WTP as a function of the usage, big usage tracking data and observed subscription choices can identify the elasticities and the distribution function of the WTP. Crucially, our approach works because purchase (subscription) is separated from usage, and the two are related in the sense that obtaining a subscription opens up for the consumer the possibility of using the service for a potentially unlimited amount. We also demonstrate how price variation, even in limited form (e.g. two price levels), can help identify the functional form of the WTP.

Even though our paper focuses on subscription markets, the idea has potential more generally. Consider markets in packaged goods which are well studied in marketing. The crucial aspect required for our method is the separation of purchase (subscription) and consumption (usage). The separation implies that consumers may have different rates of consumption after purchase. In addition, *even in typical packaged goods*, there is a separation between purchase and consumption, but in most such cases we do not observe the consumption. If consumption (usage) data were observable, our approach would be applicable to these settings too. With the advent of technological advances like 5G telecommunications and the Internet of Things, the measurement of consumption is likely to become more prevalent in the future. In fact, there are some companies that already offer such services, notably LG has a smart fridge that monitors consumption of perishables like milk with the idea that these could be automatically replenished without direct consumer intervention.[10]

A Proofs

For simplicity, we omit the subscript "i" or "(i, t)" in the proof below whenever there is no confusion.

[10] See for example: NBC News (2014)

Proof of Proposition 1. Consider the optimal quantities for the consumer $q^*(p_q, p_z, L, \phi)$ and $z^*(p_q, p_z, L, \phi)$. At the optimal, the budget constraint is binding, implying that $z^* = \dfrac{L - p_q q^*}{p_z}$ The definition of indirect utility function implies that

$$
\begin{aligned}
V &= u(\phi q^*, z^*) \\
&= q^* u\left(\phi, \frac{L}{q^*(p_q, p_z, L, \phi) p_z} - -\frac{p_q}{p_z}\right) \text{(by Homothetic property)} \\
&= q^* u\left(\phi, \frac{L}{Lf(p_q, p_z, \phi) p_z} - -\frac{p_q}{p_z}\right) \text{(by Homothetic property) } q^* = Lf(p_q, p_z, \phi) \\
&= q^* \underbrace{g(p_q, p_z, \phi)}_{\alpha_{it}}
\end{aligned}
$$

Observe the the function $g(p_q, p_z, \phi)$ depends on constant terms (price levels) and time-varying factor ϕ but not the budget L_{it}.

□

Proof of Proposition 2 and Proposition 3. First, we get the formula of $\tilde{\beta}$. Using eq. (5), we have that

$$
\begin{aligned}
S &= \mathbb{I}(\ln W \geq \ln P) \\
&= \mathbb{I}(\tilde{\beta}' \tilde{X}_1 + (\tilde{Q} - \ln P) + \eta \geq 0).
\end{aligned}
$$

Because $\tilde{Q}$ has been identified hence estimable from usage data using fixed effect estimator, $\tilde{Q} - \ln P$ serves as a "special regressor" (Lewbel, 2014). Letting

$$\hat{X}_1' = Z'[\mathrm{E}(ZZ')]^{-1}\,\mathrm{E}(Z\tilde{X}_1'),$$

it is known that (Lewbel, 2000)

$$\mathrm{E}(\hat{X}_1 \tilde{X}_1')\tilde{\beta} = \mathrm{E}(\hat{X}_1 Y),$$

hence

$$
\begin{aligned}
\tilde{\beta} &= [\mathrm{E}(\tilde{X}_1 Z')[\mathrm{E}(ZZ')]^{-1}\,\mathrm{E}(Z\tilde{X}_1')]^{-1}[\mathrm{E}(\tilde{X}_1 Z')[\mathrm{E}(ZZ')]^{-1}\,\mathrm{E}(ZY) \\
&= [\mathrm{E}(\hat{X}_1 \hat{X}_1')]^{-1}\,\mathrm{E}(\hat{X}_1 Y).
\end{aligned}
$$

Second, we obtain the formula of $F_\eta(\eta \mid \tilde{X}_1, Z)$. Let

$$\pi(x_1, z, q) = \mathrm{E}(S \mid \tilde{X}_1 = x_1, Z = z, \tilde{Q} = q).$$

We have

$$\begin{aligned}
\pi(x_1, Z, q) &= \Pr(\eta > -(\tilde{\beta}' x_1 + q - \ln P) \mid \tilde{X}_1 = x_1, Z, \tilde{Q} = q) \\
&= \Pr(\eta > -(\tilde{\beta}' x_1 + q - \ln P) \mid \tilde{X}_1 = x_1, Z) && \text{(Using Assumption 2 (ii))} \\
&= 1 - F_\eta(\ln P - \tilde{\beta}' x_1 - q \mid \tilde{X}_1 = x_1, Z).
\end{aligned}$$

In other words,

$$F_\eta(\eta \mid \tilde{X}_1, Z) = 1 - \pi(\tilde{X}_1, Z, \ln P - \tilde{\beta}' \tilde{X}_1 - \eta).$$

Third, we obtain the formula of $F_W(W \mid X, Z)$. We have

$$\begin{aligned}
F_W(w \mid X = x, Z) &= \Pr(\ln W \leq \ln w \mid X = x, Z) \\
&= \Pr(\tilde{\beta}' x_1 + \gamma_c' x_{2c} + \eta \leq \ln w \mid X = x, Z) \\
&= \Pr(\eta \leq \ln w - \tilde{\beta}' x_1 - \gamma_c' x_{2c} \mid \tilde{X}_1 = x_1, X_{2c} = x_{2c}, Z) \\
&= F_\eta(\ln w - \tilde{\beta}' x_1 - \gamma_c' x_{2c} \mid \tilde{X}_1 = x_1, X_{2c} = x_{2c}, Z) \\
&= F_\eta(\ln w - \tilde{\beta}' x_1 - \gamma_c' x_{2c} \mid \tilde{X}_1 = x_1, Z) && \text{(Using Assumption 2 (ii))} \\
&= 1 - \pi(x_1, Z, \ln P - \ln w + \gamma_c' x_{2c}).
\end{aligned}$$

Note that the above lines also establish $F_W(w \mid X, Z) = F_\eta(\ln w - \tilde{\beta}' \tilde{X}_1 - \gamma_c' X_{2c} \mid \tilde{X}_1, Z)$.

Fourth, we derive the formula of $\mathrm{E}(\ln W \mid X)$. We start with

$$\mathrm{E}(\ln W \mid X, Z) = \tilde{\beta}' \tilde{X}_1 + \gamma_c' X_{2c} + \mathrm{E}(\eta \mid X, Z).$$

When X is correlated with either U or V, $\mathrm{E}(\eta \mid X, Z) \neq 0$. By the law of iterated expectation,

$$\mathrm{E}(\ln W \mid X) = \tilde{\beta}' \tilde{X}_1 + \gamma_c' X_{2c} + \mathrm{E}(\eta \mid X).$$

Below we will show that

$$\mathrm{E}(\eta \mid X) = \mathrm{E}(H \mid X),$$

where

$$H = \mathrm{E}\left(\frac{S - \mathbb{I}(\ln P - \tilde{\beta}' \tilde{X}_1 - \tilde{Q} \leq 0)}{f_{\tilde{Q}}(\tilde{Q} \mid \tilde{X}_1, Z)} \,\middle|\, \tilde{X}_1, Z \right).$$

We first have $\mathrm{E}(\eta \mid \tilde{X}_1, X_{2c}, Z) = \mathrm{E}(\eta \mid \tilde{X}_1, Z)$ by the conditional independence assumption, thus

$$\begin{aligned}
\mathrm{E}(\eta \mid \tilde{X}_1, X_{2c}) &= \mathrm{E}(\mathrm{E}(\eta \mid \tilde{X}_1, X_{2c}, Z) \mid \tilde{X}_1, X_{2c}) \\
&= \mathrm{E}(\mathrm{E}(\eta \mid \tilde{X}_1, Z) \mid \tilde{X}_1, X_{2c}).
\end{aligned}$$

We next need to derive $\mathrm{E}(\eta \mid \tilde{X}_1, Z)$. First, it follows from the integration by parts that

$$\mathrm{E}(\eta \mid \tilde{X}_1, Z) = \int_{-\infty}^{\infty} [\mathbb{I}(\eta > 0) - F_\eta(\eta \mid \tilde{X}_1, Z)] \, \mathrm{d}\, \eta.$$

Using the identified formula of $F_\eta(\eta \mid \tilde{X}_1, Z)$, we have

$$\mathrm{E}(\eta \mid \tilde{X}_1, Z) = \int_{-\infty}^{\infty} [\mathrm{E}(S \mid \tilde{X}_1, Z, \tilde{Q} = \ln P - \tilde{\beta}' \tilde{X}_1 - \eta) - \mathbb{I}(\eta \leq 0)] \, \mathrm{d}\, \eta.$$

Using the substitution of variables, we have

$$\begin{aligned}
\mathrm{E}(\eta \mid \tilde{X}_1, Z) &= \int_{-\infty}^{\infty} [\mathrm{E}(S \mid \tilde{X}_1, Z, \tilde{Q}) - \mathbb{I}(\ln P - \tilde{\beta}' \tilde{X}_1 - \tilde{Q} \leq 0)] \, \mathrm{d}\, \tilde{Q} \\
&= \mathrm{E}\left(\frac{S - \mathbb{I}(\ln P - \tilde{\beta}' \tilde{X}_1 - \tilde{Q} \leq 0)}{f_{\tilde{Q}}(\tilde{Q} \mid \tilde{X}_1, Z)} \,\middle|\, \tilde{X}_1, Z \right) \\
&= \mathrm{E}(H \mid \tilde{X}_1, Z).
\end{aligned}$$

$\square$

Proof of Proposition 5. Based on an observation made by Lewbel (2000, page 147), we show that

$$\lambda_{it} = \eta_{it}, \tag{12}$$

where

$$\lambda_{it} = \int_{-\infty}^{\infty} \mathbb{I}(\tilde{Q}_{it,m} - \ln P > -\eta_{it}) - \mathbb{I}(\tilde{Q}_{it,m} - \ln P > 0) \, \mathrm{d}\, \tilde{Q}_{it,m}.$$

Because $U_{it} = \omega_i + \xi_t$, we then have

$$\lambda_{i,t+1} - \lambda_{it} = \xi_{t+1} - \xi_t.$$

Note that taking integral with respect ω_i both sides of the equation, we have

$$\int_\omega \lambda_{i,t+1} \, \mathrm{d}\, F(\omega) - \int_\omega \lambda_{it} \, \mathrm{d}\, F(\omega) = \xi_{t+1} - \xi_t,$$

because $\xi_t = \int_\omega \xi_t \,\mathrm{d}\, F(\omega)$. Next, we have that

$$\begin{aligned}
\int_\omega \lambda_{it} \,\mathrm{d}\, F(\omega) &= \int_\omega \int_{-\infty}^{\infty} \mathbb{I}(\tilde{Q}_{it,m} - \ln P + \xi_t > -\omega_i) - \mathbb{I}(\tilde{Q}_{it,m} - \ln P > 0) \,\mathrm{d}\, \tilde{Q}_{it,m} \,\mathrm{d}\, F(\omega) \\
&= \int_\omega \int_{-\infty}^{\infty} \frac{S_{it} - \mathbb{I}(\tilde{Q}_{it,m} - \ln P > 0)}{f_{\tilde{Q},t}(\tilde{Q}_{it,m})} f_{\tilde{Q},t}(\tilde{Q}_{it,m}) \,\mathrm{d}\, \tilde{Q}_{it,m} \,\mathrm{d}\, F(\omega) \\
&= \mathrm{E}_t\left(\frac{S_{it} - \mathbb{I}(\tilde{Q}_{it,m} - \ln P > 0)}{f_{\tilde{Q},t}(\tilde{Q}_{it,m})} \right) \\
&\equiv \mathrm{E}_t(\tilde{Y}_{it}).
\end{aligned}$$

The expectation $\mathrm{E}_t(\cdot)$ is taken with respect to the distribution of $(S_{it}, \tilde{Q}_{it,m})$ in period t only, because ξ_t is held as constant. We hence have the conclusion that

$$\xi_{t+1} - \xi_t = \mathrm{E}_{t+1}(\tilde{Y}_{i,t+1}) - \mathrm{E}_t(\tilde{Y}_{it}).$$

□

Proof of Proposition 6. Below we show that without knowing g, one cannot identify the distribution of the WTP W. To see this, let's assume that Q^* is observed, and let's ignore $\tilde{X}_1$ and Z, and let $\ln \alpha = U$. We have

$$S = \mathbb{I}(W > P) = \mathbb{I}(g(U + \ln Q^*) > P) = \mathbb{I}(U + \ln Q^* > g^{-1}(P))$$

in this simple case.

Observe that

$$\begin{aligned}
\mathrm{E}(S \mid \ln Q^* = q) &= \Pr(U > -q + g^{-1}(P) \mid \ln Q^* = q) \\
&= 1 - F_U(g^{-1}(P) - q).
\end{aligned}$$

Hence we obtain the CDF of U_{it} as follows,

$$F_U(u; g) = 1 - \mathrm{E}(S \mid \ln Q^* = g^{-1}(P) - u),$$

and the conditional CDF of the WTP W given Q^* as follows,

$$F_W(w \mid Q^* = q; g) = 1 - \mathrm{E}(S \mid \ln Q^* = g^{-1}(P) - g^{-1}(w) + \ln q).$$

So the identified $F_U(U; g)$ and $F_W(W \mid Q^*; g)$ will change with respect to g. □

References

Becker, G. M., M. H. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Belomestnyi, D. (2002). On the problem of characterizing the distribution of random variables by the distribution of their sum. *Journal of Mathematical Sciences 111*(3), 3498–3504.

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

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

Breidert, C. (2007). *Estimation of Willingness-To-Pay: Theory, Measurement, Application.* Springer Science & Business Media.

Chen, T., K. Fenyo, S. Yang, and J. Zhang (2018). Thinking inside the subscription box: New research on e-commerce consumers. *McKinsey Quarterly*. Accessed: 2019-06-20.

Columbus, L. (2018). The state of the subscription economy, 2018. *Forbes*. Accessed: 2019-06-20.

Danthurebandara, V. M., J. Yu, and M. Vandebroek (2011). Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. *Quantitative Marketing and Economics 9*(4), 429–448.

Diamond, P. A. and J. A. Hausman (1994). Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives 8*(4), 45–64.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research 44*(2), 214–223.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review 97*(3), 713–744.

Green, P. E. and V. R. Rao (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing research 8*(3), 355–363.

Green, P. E. and V. Srinivasan (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research 5*(2), 103–123.

Guadagni, P. M. and J. D. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science 2*(3), 203–238.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives 8*(4), 19–43.

Hausman, J. (2012). Contingent valuation: From dubious to hopeless. *Journal of Economic Perspectives 26*(4), 43–56.

Heckman, J. J. (1979). Sample selection as a specification error. *Econometrica 47*, 153–161.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010). Nonparametric identification and estimation of nonadditive hedonic models. *Econometrica 78*(5), 1569–1591.

Kalish, S. and P. Nelson (1991). A comparison of ranking, rating and reservation price measurement in conjoint analysis. *Marketing Letters 2*(4), 327–335.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A. (2014). An overview of the special regressor method. In J. S. Racine, L. Su, and A. Ullah (Eds.), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Chapter 2, pp. 38–64. Oxford University Press.

Lewbel, A., D. McFadden, and O. Linton (2011). Estimating features of a distribution from binomial data. *Journal of Econometrics 162*(2), 170–188.

Mitchell, R. C. and R. T. Carson (2013). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Rff Press.

NBC News (2014). Out of milk? lg's new smart fridge will let you know. https://www.nbcnews.com/tech/gift-guide/out-milk-lgs-new-smart-fridge-will-let-you-know-n99531. Accessed: 2019-06-20.

Nevo, A., J. L. Turner, and J. W. Williams (2016). Usage-based pricing and demand for residential broadband. *Econometrica 84*(2), 411–443.

Noussair, C., S. Robin, and B. Ruffieux (2004). Revealing consumers' willingness-to-pay: A comparison of the bdm mechanism and the vickrey auction. *Journal of Economic Psychology 25*(6), 725–741.

Rao, V. R. (2014). *Applied Conjoint Analysis*. Springer.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy 82*(1), 34–55.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15*(4), 321–340.

Shi, R. (2019). Identification and estimation of nonparametric hedonic equilibrium model with unobserved quality. Working paper, UC Riverside.

Train, K. and M. Weeks (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 1–16. Springer.

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy 97*(2), 444–479.

Villas-Boas, J. M. and R. S. Winer (1999). Endogeneity in brand choice models. *Management Science 45*(10), 1324–1338.

Voelckner, F. (2006). An empirical comparison of methods for measuring consumers' willingness to pay. *Marketing Letters 17*(2), 137–149.

Wertenbroch, K. and B. Skiera (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research 39*(2), 228–241.

Xie, J. and S. M. Shugan (2001). Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Science 20*(3), 219–243.