Can Willingness to Pay be Identified without Price Variation?
What Big Data on Usage Tracking Can (and Cannot) Tell Us

Cheng Chou Vineet Kumar*

Fall 2019

Abstract
We study how to obtain the distribution of consumers’ willingness to pay (WTP) for subscription products, where consumers pay a monthly fixed price. We show that variation in usage and subscription choice together can identify the elasticities and the WTP distribution. We propose a novel estimation strategy to recover the WTP distribution in the absence of price variation. In addition, we show how price variation can help identify the functional form in which the usage affects the WTP. We demonstrate an application of our method with both usage and monthly subscription data from a music streaming service. We recover the conditional distribution of WTP based on demographic variables, and find negative age elasticity of the usage, but positive age elasticity of the WTP, and that male subscribers are willing to pay more for the service. We estimate the demand curve for the monthly plan and the distribution of WTP for a counterfactual subscription plan, and we determine how to optimally set prices for the new plan.

1 Introduction

Subscriptions are becoming increasingly popular across the world for both physical and digital products and services with growth over 100% over 2013–2018 (Columbus [2018] Chen,}

*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 also thank participants in seminars at University of Sheffield, UCL and Yale University.
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. There are a number of reasons for this popularity, including low marginal costs (relative to fixed costs), reducing consumer risk, avoiding transaction costs, providing predictability in revenue stream as well as increasing loyalty (Xie and Shugan, 2001).

Our paper studies how to obtain the distribution of consumers willingness to pay (WTP) for subscription products. Estimating the distribution of WTP, given consumer and product characteristics is a first step required to understand and predict demand responses, to identify how consumers value various features of the product or service, and to decide how alternative products should be priced. Consider the example of Netflix, which has a monthly plan for $8.99 in the US. When the firm is interested in evaluating how demand might vary with price increases, or identifying how a counterfactual weekly plan should be priced, we would need to obtain the WTP distribution.

In most subscription markets, price variation is fairly rare or non-existent, except for free trials. The absence of price variation presents the major challenge in identifying the distribution of WTP, because 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 the WTP, which relies on price variation, a feature common to the entire literature (Guadagni and Little, 1983; Dantherebandara, Yu and Vandebroek, 2011; Lewbel, McFadden and Linton, 2011; Train and Weeks, 2005). Firms in such markets set these 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 capturing accurately 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, the approach uses variation of usage (or consumption) of the subscription. Below we will use “usage” and “consumption” interchangeably. In subscription models, purchase decisions are relatively low-frequency with consumers choosing among plans every month or so, whereas usage is typically high-frequency, with possibly daily or intra-day data often available. Thus, the “big data” in the title of the paper refers to high-frequency usage.

1As 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

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

data. Overall, the typical data available in subscription settings includes product usage data, subscription/churn choices, and often rich data on a variety of consumer and product characteristics.

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 big data on usage? Second, in settings with price variation, what additional inference is possible? Third, what classes of counterfactuals are identified.

The main contribution of this paper is to propose a novel method to identify and estimate semiparametrically the conditional distribution of the WTP given product features and customer characteristics **when price variation is absent**. 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, e.g. Netflix has plans at $8.99, $12.99 and $15.99 per
month. Rather, it works with only one plan present, e.g. Apple Music only has one plan at $9.99 per month.

The second contribution is that we demonstrate how to use the estimated conditional distribution WTP to guide a wide variety of product and pricing choices. First, the firm might want to assess the impact of improvements in product quality or other product characteristics on the distribution of consumer valuation. Second, a firm might consider changing the subscription period. For example, introducing a weekly plan or an annual plan in place of a monthly plan. The firm might also evaluate introducing quotas or quantity restrictions. Such restrictions are often imposed in otherwise unlimited plans for mobile data service and cable plans. Netflix and Comcast have been reported to throttle some heavy users.\footnote{See for example: Associated Press (2006)}

Broadly, our findings point to both the potential and limitation of usage data in recovering consumers 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 WTP for one unit of usage on average. 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 for one unit of usage, as well as the \textit{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 for a unit of usage relies on the following argument. Even though price is identical among consumers, there is still variation in the “price” of \textit{per-unit} usage when there is variation of usage among consumers. We then can recover the distribution of the WTP for one unit of 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 identify the joint distribution of expected usage and WTP for a unit of usage.

We take our method to data using an application of music streaming, featuring monthly subscription and daily 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. Finally, we estimate the demand curve for the monthly plan and the obtain distribution of the WTP for a counterfactual weekly subscription plan, allowing us to assess its revenue impact.

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 the Conclusion how it can be applied to say, packaged goods.

Even though our paper only focuses on subscription markets, the idea has potential more generally to even say, packaged goods. The crucial aspect required for our method is the separation of purchase (subscription) and consumption (usage). The separation implies that two consumers typically have different amount of consumption if they have both purchased. Observe that this holds naturally in subscription settings like those in Table 1. In addition, *even in typical packaged goods*, there is a separation between purchase and consumption, but on most such cases we do not observe the consumption. If consumption (usage) data were observable, our approach would be applicable to such settings too. With the advent of technological advances like the 5G telecommunications and the Internet of Things, the measurement of such 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.

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 6 studies the role of price variation in identifying the WTP distribution. Section 7 uses the approach in an empirical application of music streaming to demonstrate its potential value. The appendices contain technical proofs and Monte Carlo studies of the performance of our method.

## 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

---

3See for example: NBC News (2014)
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 remains used in contingent valuation type settings without product variation. For example, consumers might be asked how much they value a particular public and environmental goods like 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 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 the researcher (or firm) to introduce exogenous variation in product characteristics or attributes, prices as well as in 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 market-level (Berry 1994; Berry, Levinsohn and Pakes 1995) or micro-data based on individual consumers like in much of the marketing literature (Guadagni and Little 1983). In 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 the 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.

Perhaps the most 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. Under this three-part tariff, subscribers pay a fixed fee each month that pays for all usage up to a certain allowance, and they will be charged at an overage price for each GB of
usage in excess of the 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. The key difference between the present paper and Nevo et al. (2016) is that our identification arguments do not rely on the presence of three-part tariff with overage rates that creates the variation of shadow price of usage as the accumulated usage approaches to allowance. This is relevant 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 Essential Logic of Identification

To develop intuition, we provide a model of consumer purchase and usage in a subscription market, featuring several simplifying assumptions. Consider a consumer who purchases an unlimited monthly gym membership at price $P$ that does not vary across consumers or over time. We are interested in obtaining the distribution of WTP for the gym’s customer base (i.e. those who have purchased the service some time in the past).

Let $S_i \in \{0 \text{ (no)}, 1 \text{ (yes)}\}$ denote the observable subscription choice—$S_i = 1$ if individual $i$ have the subscription. We adopt a money-metric representation of expected utility of the service. Let

$$u_i = W_i - P$$

be the expected utility of the service, if $i$ subscribed. Here $W_i$ is her WTP. If she did not subscribe, there is an outside option (e.g., running) for her, and the money-metric expected utility of this outside option equals $\mu = 0$.

She will subscribe if and only if she believes that the expected utility of the service, $W_i - P$, is greater than the expected utility of the outside option.

Consider the customer base of a gym that charges a monthly membership fee of $30, having two segments of customers A and B. Segment A comprise 40% of the population and visit the gym 10 times per month. Segment B comprise 60% and visit the gym 5 times per month. Thus, the effective price per unit usage (visit) is different for these segments. The segments also have different churn or retention rates with segment A customers retained at 90% and B at 80%.

In Section XYZ of the paper, we demonstrate under what conditions $\mu$ is identified.
Table 2: Gym Example

<table>
<thead>
<tr>
<th>Segment</th>
<th>Proportion</th>
<th># Gym Visits Per Month</th>
<th>Monthly Fee</th>
<th>&quot;Price&quot; Per Visit</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>40%</td>
<td>10</td>
<td>$30</td>
<td>$3</td>
<td>90%</td>
</tr>
<tr>
<td>B</td>
<td>60%</td>
<td>5</td>
<td>$30</td>
<td>$6</td>
<td>80%</td>
</tr>
</tbody>
</table>

We can obtain the overall WTP distribution of the customer base as follows:

\[
P(WTP \text{ for Gym Membership} < \$30) = 0.4 \cdot P(WTP \text{ per visit for A} < \$3) + 0.6 \cdot P(WTP \text{ per visit for B} < \$6)
\]

\[
= 0.4 \times (1 - 0.9) + 0.6 \times (1 - 0.8) = 0.16
\]

More generally, let a consumer \( i \) have an expected number of visits \( Q_i^* \). Denote her average WTP for a visit by \( \alpha_i \). Her WTP for the membership is then \( W_i = \alpha_i Q_i^* \). Note that the above does not imply that she has a linear valuation for usage.

Suppose \( Q^* \) and \( \alpha \) are independent, which is not an innocuous assumption. Denote the CDF of \( \alpha \) by \( F_\alpha \). Observe that the subscription retention decision for the consumer is:

\[
S_i = \mathbb{I}(W_i > P) = \mathbb{I}(\alpha_i Q_i^* > P) = \mathbb{I}(\alpha_i > P/Q_i^*)
\]

We then get the probability of retention, conditional on usage to be:

\[
P(S_i = 1 | Q_i^* = q) = P(\alpha_i > P/q) = 1 - F_\alpha(P/q)
\]

When \( Q_i^* \) is observable, then the object \( P(S_i = 1 | Q_i^* = q) \) is directly observable from data. Then, if \( Q_i^* \) has large enough variation, \( P/Q_i^* \) will cover the support of \( \alpha_i \), and \( F_\alpha \) can be identified. Observe that if we do not have any variation in usage, i.e. \( Q_i^* \) is a constant, then we only know the value of \( F_\alpha \) at one point, and the distribution is not identified. This highlights the role of usage variation in helping obtain WTP.

It is worthwhile explaining why we have assumed independence of \( \alpha_i \) and \( Q_i^* \). When \( Q_i^* \) is not independent of \( \alpha_i \), we can only identify the particular probabilities \( F_\alpha(P/q \mid Q_i^* = q) \) for different values of \( Q_i^* \) rather than the entire distribution \( F_\alpha(\cdot \mid Q_i^* = q) \). We do not identify \( F_\alpha(P/q' \mid Q_i^* = q) = \Pr(\alpha_i \leq P/q' \mid Q_i^* = q) \) for any \( q' \neq q \).

The above example illustrates how the two sources of variation (retention and usage) allow us to obtain WTP for the service. The key observation is that for a subscription-like service, the purchasing and usage (consumption) are separated in the sense that in data we can observe two subscribers who have different amount of usage, but paid the same price
for the plan. As a result, though the price is constant for the subscribed plan, there is still variation in the price per unit usage. More broadly, this variation in price per unit ex-post usage can either be across consumers like in Table 2 or within a consumer across time periods.

The above argument makes a number of simplifying assumptions, which we seek to relax in the rest of the paper. First, there is typically a continuum of heterogeneous consumers rather than the two segments considered above. Moreover, such consumers may be heterogeneous (even within a segment) in both usage and WTP per unit of usage (i.e. WTP per visit in the Gym example). Second, and perhaps most important, we have assumed above that WTP per usage unit ($\alpha_i$) and the expected usage ($Q^*_i$) are independent, which is unlikely to hold in practice. Third, we have not considered any observed consumer or product characteristics that might impact WTP. Fourth, the recovered WTP is only valid for the customer base for which we have the observations. Extending this to the population of consumers requires us to deal with selection issues, which results when consumers with low unobserved preferences for gym membership are not part of the customer base, from which we make inferences.

Below, we detail our general framework that accommodates each of the above issues.

4 Model

We now detail the main model. Here, we relax the major simplifying assumptions that we had made above for expositional simplicity. We observe panel data about $n$ number of consumers over $T$ subscription decisions. We use panel data to deal with the complications created by the unobserved consumer heterogeneity in WTP for one unit of usage and expected usage, as well as selection issues mentioned earlier.

Let $W_{it}$ be consumer $i$’s WTP for the subscription service in billing period $t$, and let $Q^*_{it}$ be her expected usage of the service in the billing period $t$. The other notation such as $\alpha_{it}$, $X_{1it}$, and $U_{it}$ are defined similarly. We still maintain the assumption that the expected usage $Q^*_{it}$ is observed in this subsection.

4.1 Unobserved Heterogeneity

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. We let $\omega_i$ and $\xi_t$ to denote the former and the latter, respectively, and let $U_{it} = \omega_i + \xi_t$. Below, we
are going to show that with panel data we indeed can separately identify the distributions of $\omega_i$ and $\xi_t$ under certain conditions. To simplify the exposition, we omit $X_{1it}$ and assume $Q_{it}^{*} \perp \perp U_{it}$ below.

Consider the linear projection of $\ln \alpha_i$ on the vector of observable covariates $X_{1it}$,

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

where $U_{it}$ denotes the unobservable heterogeneity that affects consumer $i$’s WTP for one unit of usage. The first element of $X_{1it}$ is unity with coefficient $\beta_1$, and $\mathbb{E}(U_i) = 0$.

When we do not observe any $X_{1it}$, $\alpha_{it} = e^{\beta_1 + U_{it}}$ is just a random coefficient of expected usage, which can be interpreted as a demand shifter. When $X_{1it}$ is “endogenous”, that is $X_{1it}$ and $U_{it}$ are correlated, we need a vector of instrumental variables (IV) $Z_{it}$ that is uncorrelated with $U_{it}$.

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. We let $\omega_i$ and $\xi_t$ to denote the former and the latter, respectively, and let $U_{it} = \omega_i + \xi_t$. Below, we are going to show that with panel data we indeed can separately identify the distributions of $\omega_i$ and $\xi_t$ under certain conditions. To simplify the exposition, we omit $X_{1it}$ and assume $Q_{it}^{*} \perp \perp U_{it}$ below.

The idea is to note that $\omega_i$ 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. (2) below). If one is willing to assume an initial value $\xi_0$, then one identifies each value of $\xi_t$, hence the distribution of $\xi_t$. Then the distribution of $\omega_i$ is easily obtained from the distribution of $U_{it} - \xi_t$ since we have identified the distribution of $U_{it}$ using ?? As an alternative, one can assume that $\xi_t$ 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 $\omega_i$ by deconvolution of the distribution of $U_{it} = \omega_i + \xi_t$ from assuming that $\xi_t \perp \perp \omega_i$.

Below, we provide a formula of $\xi_{t+1} - \xi_t$. Based on an observation made by Lewbel (2000, page 147), we show in the appendix that

$$\lambda_{it} = U_{it},$$

where

$$\lambda_{it} = \int_{-\infty}^{\infty} \mathbb{I}(\ln Q_{it}^{*} - \ln P > -U_{it}) - \mathbb{I}(\ln Q_{it}^{*} - \ln P > 0) \ d(\ln Q_{it}^{*}).$$

11
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 $\omega_i$ both sides of the equation, we have

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

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

$$\int_{\omega} \lambda_{it} dF(\omega) = \int_{\omega} \int_{-\infty}^{\infty} \mathbb{I}(\ln Q_{it}^* - \ln P > \omega_i) - \mathbb{I}(\ln Q_{it}^* - \ln P > 0) dQ_{it}^* dF(\omega)$$

$$= \int_{\omega} \int_{-\infty}^{\infty} \frac{S_{it} - \mathbb{I}(\ln Q_{it}^* - \ln P > 0)}{f_{\ln Q,t}(Q_{it}^*)} f_{\ln Q,t}(Q_{it}^*) dQ_{it}^* dF(\omega)$$

$$= \mathbb{E}_t \left( \frac{S_{it} - \mathbb{I}(\ln Q_{it}^* - \ln P > 0)}{f_{\ln Q,t}(Q_{it}^*)} \right)$$

$$\equiv \mathbb{E}_t(\tilde{Y}_{1it}).$$

Here $f_{\ln Q,t}(Q_{it}^*)$ is the density function of $\ln Q_{it}^*$ in period $t$. The expectation $\mathbb{E}_t(\cdot)$ is taken with respect to the distribution of $(S_{it}, Q_{it}^*)$ in period $t$ only, because $\xi_t$ is held as constant. We hence have the conclusion that

$$\xi_{t+1} - \xi_t = \mathbb{E}_{t+1}(\tilde{Y}_{1i,t+1}) - \mathbb{E}_t(\tilde{Y}_{1it}).$$  \hfill (2)

Obviously, the right-hand-side (RHS) is estimable from data by sample averages.

### 4.2 Is Expected Usage Known?

In the study of churn decisions ($S_{it} = 0$ if $i$ churn before the beginning of period $t$, and $S_{it} = 1$ otherwise), the sample contains the historical usages of all consumers, who have ever subscribed to the service. Let $Q_{it}$ be the observed consumption in billing period $t$.

**Assumption 1** (Expected usage and observed usage: I). Assume $\ln Q_{it}^* = \ln Q_{i,t-1} + \zeta_{it}$, and $\zeta_{it} \perp Q_{i,t-1}$.\footnote{A special case is that $\text{Var}(\zeta_{it}) = 0$, and consumers determine their expected usage based on the usage in the previous month.}

The subscription choice can then be written in terms of the actual usage as follows,

$$S_{it} = \mathbb{I}(\ln W_{it} > \ln P) = \mathbb{I}(\beta' X_{1it} + \ln Q_{i,t-1} - \ln P + (U_{it} + \zeta_{it}) > 0).$$  \hfill (3)
4.3 Extending to the Population of Consumers

In most applications, we do not observe the usage by non-subscribers. Below, we posit that although we cannot observe the usage by each individual, we observe consumer and product characteristics \( X_{2it} \) for the population, which in turn impacts usage.

Using the data about subscribers whose usage is observed, we project the usage onto the space of \( X_{2it} \), then impute the “usage” driven by the variation of \( X_{2it} \) for each individual. Sample selection (subscribers and non-subscribers are different in unobserved characteristics that affect usage) is not a hard issue, when we have panel data about usage and characteristics \( X_{2it} \). In the gym example, \( X_{2it} \) can include the distance to the gym, leisure time and weather.

Assumption 2 (Expected Usage and Selection). (i) Assume that the log of expected consumption \( \ln Q_{it}^\ast \) has the following reduced form regardless of the subscription choice \( S_{it} \),

\[
\ln Q_{it}^\ast = \gamma'X_{2it} + V_i,
\]

where the unobserved fixed effect \( V_i \) can be correlated with \( U_{it} \) in the specification of the WTP for one unit of usage \( (\alpha_i) \) to allow for selection. Assume that \( \text{E}(V_i) = 0 \).

(ii) Let \( Q_{it} \) be the observed consumption when \( S_{it} = 1 \), and assume that

\[
\ln Q_{it} = \ln Q_{it}^\ast + \varepsilon_{it},
\]

where \( \varepsilon_{it} \) is serially uncorrelated random shock.

(iii) Assume strict exogeneity \( \text{E}(\varepsilon_{it} \mid X_i, U_i, V_i) = 0 \), where \( X_i = (X_{i1}^t, \ldots, X_{iT}^t)' \) and \( U_i = (U_{i1}, \ldots, U_{iT})' \). Here \( X_{it} \) is the union of \( X_{1it} \) and \( X_{2it} \).

(iv) All covariates \( X_{2it} \) are time varying, and \( X_{2it} \) are observable for both subscribers and non-subscribers.

Assumption 2 (iv) is of course restrictive (e.g., by assuming all elements of \( X_{2it} \) are time varying, \( X_{2it} \) does not even include intercept term). In Remark 1, one will see that how to extend the analysis to allow for time invariant covariates.

Let \( S_i = (S_{i1}, \ldots, S_{iT})' \) be the vector of subscription choices in the \( T \) sampling periods. Because \( S_i \) is a function of \( X_i, U_i, V_i \), we have

\[
\text{E}(\varepsilon_{it} \mid X_i, V_i, S_i) = \text{E}(\text{E}(\varepsilon_{it} \mid X_i, U_i, V_i) \mid X_i, V_i, S_i) = 0.
\]

Let \( \ln Q_i = T^{-1} \sum_{t=1}^{T} \ln Q_{it} \) and \( \bar{X}_{2it} = T^{-1} \sum_{t=1}^{T} X_{2it} \). Using \( \varepsilon_{it} = \ln Q_{it} - \gamma'X_{2it} - V_i \), we have

\[
\text{E}(\ln Q_{it} - \gamma'X_{2it} - V_i \mid X_i, V_i, S_i = 1) = 0,
\]

and

\[
\text{E}(\ln Q_{it} - \ln Q_i - \gamma'(X_{2it} - \bar{X}_{2i}) \mid X_i, V_i, S_i = 1) = 0,
\]
from which the $\gamma$ associated with time varying $X_{2it}$ can be identified and estimated by the usual fixed effect (FE) estimator using the selected sample ($S_{it} = 1$). Of course FE estimator cannot identify the $\gamma$ associated with time invariant $X_{2it}$.

Let
\[ \tilde{Q}_{it} \equiv \gamma'X_{2it} \quad \text{and} \quad \eta_{it} \equiv U_{it} + V_i. \]

We have $\ln Q_{it}^* = \tilde{Q}_{it} + V_i$, and
\[ S_{it} = I(\beta'X_{1it} + \tilde{Q}_{it} - \ln P + \eta_{it} > 0). \]

Because $\tilde{Q}_{it}$ is identified for subscribers and non-subscribers, we can view $\tilde{Q}_{it}$ as a special regressor and apply similar arguments in the last section to the above display. Note that $\tilde{Q}_{it}$ should be only understood as the log of the identified consumption component that is driven only by $X_{2it}$.

**Assumption 3.**

(i) *(Big support of consumption).* The support of $\tilde{Q}_{it} \mid (X_{1it}, Z_{it})$ covers the support of $\ln(P/\alpha_{it}) - V_i \mid (X_{1it}, Z_{it})$;

(ii) *(Independence of consumption).* $X_{2it} \perp \perp (U_{it}, V_i) \mid (X_{1it}, Z_{it})$;

(iii) *(Valid and relevant IV).* $E(Z_{it}\eta_{it}) = 0$, $E(Z_{it}Z_{it}')$ is nonsingular, and $\text{rank } E(X_{1it}Z_{it}') = \dim(X_{1it})$.

(iv) *(Boundary conditions).* The support of $\eta_{it}$ contains 0, and $\lim_{\eta \to \infty} \eta[F_{\eta}(\eta \mid X_{1it}, Z_{i}) - 1] = \lim_{\eta \to -\infty} \eta F_{\eta}(\eta \mid X_{1it}, Z_{i}) = 0$.

Assumption 3 (i) and 3 (ii) imply that $X_{2it}$ must have excluded variables that affect usage but do not affect the valuation of one unit of usage (hence does not belong to $X_{1it}$), and such excluded variables are independent of $U_{it}$ and $V_i$ given $X_{1it}$ and $Z_{it}$. If $X_{2it}$ is a subset of $X_{1it}$, the support condition of Assumption 3 (i) fails as the support of $\tilde{Q}_{it} = \gamma'X_{2it} \mid (X_{1it}, Z_{it})$ includes only one point. In gym usage example, such an excluded variable can be weather in the residence of consumer $i$. In music streaming example, such excluded variables could be the measurement about the access to internet (e.g., WiFi or 4G network coverage, quota of cellular data). The access to internet affects the usage but do not affect one’s valuation of one unit of usage, say one song, after controlling variables like income.

**Proposition 1** *(Elasticities of the WTP: unobserved expected usage).* Suppose ??, 3 and 3 hold. We have
\[ E(\ln W_{it} \mid X_{1it}, X_{2it}) = \beta'X_{1it} + \gamma'X_{2it} + E(\eta_{it} \mid X_{1it}, X_{2it}), \]
where \( E(\eta_{it} \mid X_{1it}, X_{2it}) = E(H_{2it} \mid X_{1it}, X_{2it}) \) with
\[
H_{2it} = E\left( \frac{S_{it} - \mathbb{I} (\ln P - \beta' X_{1it} - \tilde{Q}_{it} \leq 0)}{f_{\tilde{Q}}(\tilde{Q}_{it} \mid X_{1it}, Z_{it})} \right) \mid X_{1it}, Z_{it} \right).
\]

Here \( f_{\tilde{Q}}(\tilde{Q}_{it} \mid X_{1it}, Z_{it}) \) is the PDF of \( \tilde{Q}_{it} \) given \( (X_{1it}, Z_{it}) \).

We have the following 2SLS formula for \( \beta \),
\[
\beta = [E(\tilde{X}_{1it}\tilde{X}'_{1it})]^{-1} E(\tilde{X}_{1it}Y_{2it}),
\]
where \( \tilde{X}'_{1it} = Z'_i[E(Z_{it}Z'_{it})]^{-1} E(Z_{it}X'_{1it}) \), and
\[
Y_{2it} = \frac{S_{it} - \mathbb{I}(\hat{Q}_{it} - \ln P > 0)}{f_{\hat{Q}}(\hat{Q}_{it} \mid X_{1it}, Z_{it})}.
\]

\( \beta \) is the probability limit of the 2SLS estimator of regressing \( Y_{2it} \) on \( X_{1it} \) using IV \( Z_{it} \).

We now have a formula of \( E(\ln W_{it} \mid X_{1it}, X_{2it}) \), which can be used to calculate the elasticities of the WTP to \( X_{1it} \) and \( X_{2it} \). When \( X_{1it} \) and \( X_{2it} \) are mean independent of \( \eta_{it} \), the elasticities of the WTP depends only on \( \beta \) and \( \gamma \). Using the usage data, one can estimate \( \gamma \) and \( \beta \) by a 2SLS regression. The following steps outline the procedure.

**Step 1:** Fixed effect estimator for log-usage. Using the data of the subscribers, estimate the following panel data model
\[
\ln Q_{it} = \gamma'X_{2it} + V_i + \varepsilon_{it},
\]
with fixed effect estimator. Let \( \hat{\gamma} \) be the estimator of \( \gamma \), and let \( \hat{Q}_{it} = \hat{\gamma}'X_{2it} \) for all \( i \) (subscribers and non-subscribers) and \( t \). When \( X_{2it} \) includes the unity term and other time invariant covariates, the associated parameters are not identified. We can apply the arguments of Remark 1.

**Step 2:** Nonparametric estimation of \( f_{\hat{Q}}(\hat{Q}_{it} \mid X_{1it}, Z_{it}) \). Obtain a nonparametric estimator \( \hat{f}_{\hat{Q}}(\cdot \mid X_{1it}, Z_{it}) \) of the conditional PDF \( f_{\hat{Q}}(\hat{Q}_{it} \mid X_{1it}, Z_{it}) \) from the sample \( (\hat{Q}_{it}, X_{1it}, Z_{it}) \) for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \).

**Step 3:** IV regression to estimate \( \beta \). Estimate \( \beta \) by the 2SLS IV regression of \( \hat{Y}_{2it} \) on \( X_{1it} \) using IV \( Z_{it} \), where
\[
\hat{Y}_{2it} = \frac{S_{it} - \mathbb{I}(\hat{Q}_{it} - \ln P > 0)}{\hat{f}_{\hat{Q}}(\hat{Q}_{it} \mid X_{1it}, Z_{it})}.
\]

**Proposition 2** (CDF of the WTP: unobserved expected usage). Suppose \( \lambda \), \( \beta \) and \( \gamma \) hold. Define the CCP function, \( \pi_2(X_{1it}, Z_{it}, \tilde{Q}_{it}) = E(S_{it} \mid X_{1it}, Z_{it}, \tilde{Q}_{it}) \). We have
\[
F_\eta(\eta \mid X_{1it}, Z_{it}) = 1 - \pi_2(X_{1it}, Z_{it}, \ln P - \beta' X_{1it} - \eta),
\]
\[
F_W(w \mid X_{it}, Z_{it}) = 1 - \pi_2(X_{1it}, Z_{it}, \ln P - \ln w + \gamma' X_{2it}),
\]
and \( F_W(w \mid X_{it}, Z_{it}) = F_\eta(\ln w - \beta' X_{1it} - \gamma' X_{2it} \mid X_{1it}, Z_{it}) \).

15
Remark 1 (What are really $X_{1it}$ and $X_{2it}$?). We now show how to relax Assumption 2 (iv) and Assumption 3 (ii) by allowing $X_{2it}$ to include time invariant covariates and/or to include covariates that are correlated with either $U_{it}$ or $V_i$. The most obvious example is that $X_{2it}$ includes the intercept (unity) term. Suppose

$$\ln Q_{it}^* = \gamma'_a X_{2it,a} + \gamma'_b X_{2it,b} + V_i,$$

where $X_{2it,b}$ is the vector of time varying covariates, and the vector $X_{2it,a}$ includes time invariant covariates (gender, location) or the covariates that are not independent of $(U_{it}, V_i)$ given $(X_{1it}, Z_{it})$. By the fixed effect estimator, we can still identify $\gamma_b$, but not necessarily $\gamma_a$. Under the above specification, we have

$$S_{it} = I((\beta'X_{1it} + \gamma'_a X_{2it,a}) + \gamma'_b X_{2it,b} - \ln P + \eta_{it} > 0).$$ (4)

Observe that it is similar to the model

$$S_{it} = I(\beta'X_{1it} + \gamma'X_{2it} - \ln P + \eta_{it} > 0).$$ (5)

We can redefine $\beta$, $X_{1it}$ and $X_{2it}$ in eq. (5) so that $\beta'X_{1it} + \gamma'_a X_{2it,a}$ of eq. (4) becomes $\beta'X_{1it}$ of eq. (5), and $\gamma'_b X_{2it,b}$ of eq. (4) becomes $\gamma'X_{2it}$ of eq. (5), that is also $\bar{Q}_{it}$.

When the support of $\gamma'_b X_{2it,b}|(X_{1it}, X_{2it,a}, Z_{it})$ covers the support $(\ln P - \beta'X_{1it} - \gamma'_a X_{2it,a} - \eta_{it}) | (X_{1it}, X_{2it,a}, Z_{it}), X_{2it,b} \perp (U_{it}, V_i) | (X_{1it}, X_{2it,a}, Z_{it})$, and there are enough number of IVs, we can apply Proposition 1 and Proposition 2 to identify $F_W(w | X_{it}, Z_{it})$ and $E(\ln W_{it} | X_{it}, Z_{it})$.

The caveat is that if $X_{1it}$ and $X_{2it,a}$ has intersections (e.g., when both $X_{1it}$ and $X_{2it,a}$ have unity term), and $\gamma_a$ cannot be identified from consumption data (e.g., due to time invariant nature of such variables), we cannot separately identify $\beta$ and $\gamma_a$ from the subscription choice. Instead, we can only identify $\beta + \gamma_a$.

5 Applications

5.1 Pricing a shorter subscription service based on higher frequency usage data

Usually consumption has seasonality and serial correlation. When the distribution of consumption of a subscription plan varies within its billing period (e.g., yearly plan), consumers may want a plan of shorter length (e.g., monthly plan) or a plan for a particular period (holiday season). With the access to high frequency consumption data (e.g., daily consumption
data), we show that it is possible to identify the distribution of the WTP for plans of varying lengths or plans for certain specific periods.

A reasonable question here is why we need a model to set pricing for different lengths of plan. For example, why not use a simple approach of just dividing the price by the period length, for example a monthly subscription should be priced at $1/12$ of an annual subscription, since the mean monthly usage is on average $1/12$ of annual usage? This argument does not hold, since we need the distribution of monthly usage to be exactly the distribution of $1/12$ of the annual usage for this to work. When there is serial correlation in usage across different periods, this will not hold. For example, suppose the 12 monthly usage variables have identical distributions, the variance of $1/12$ of the annual usage is always smaller than the variance of monthly usage unless the 12 monthly usages are perfectly correlated. Thus, in general the optimal price for a monthly plan may be different from that of scaled down annual price.

Another reason for varied valuation over time periods is that peak and off-peak usage might be different. To fix the idea, suppose the subscription period is one year, and it consists of peak and off-peak periods. Consumers use the plan more during the peak period. Also, assume that $X_{it}$ are exogenous, so $Z_{it} = X_{it}$. Let $t$ denote one year, and let $t_p$ and $t_{op}$ denote the peak and off-peak periods in year $t$. Similarly, let $Q_{it_p}$ and $Q_{it_{op}}$ be the observed consumption in the two periods, and $Q_{it} = Q_{it_p} + Q_{it_{op}}$.

From the actual yearly consumption $Q_{it}$, we can first identify the distribution $F_{\eta}(\eta_{it}|X_{1it})$. We then can identify the distribution function of the WTP for plan during peak and off-peak period. Taking the peak period for example, the log of the WTP for the plan of the peak period is

$$\ln W_{it_p} = \ln \alpha_{it_p} + \tilde{Q}_{it_p} + V_i = \beta' X_{1it_p} + \gamma_p' X_{2it_p} + V_i + U_{it_p},$$

where $\gamma_p$ can be identified from the consumption in peak period. If we assume that $U_{it_p} = U_{it}$, we have $V_i + U_{it_p}$ and $\eta_{it}$ have identical distribution function $F_{\eta}(\eta|X_{1it})$. We then can identify the distribution function of $W_{it_p}$:

$$F_{W_{it_p}}(w \mid X_{1it} = x_1, X_{2it} = x_2) = F_{\eta}(\ln w - \beta' x_1 - \gamma_p' x_2 \mid X_{1it} = x_1).$$

Because the assumption $U_{it_p} = U_{it}$ is important, it is instructive to provide examples, in which this assumption holds and fails. The interpretation of $U_{it}$ or $U_{it_p}$ depends on what have been controlled by $X_{1it}$ or $X_{1it_p}$. In the gym subscription example, first consider the following specifications

$$\ln \alpha_{it} = \beta_1 + \beta_2 GymAval_{it} + \beta_3 INC_{it}, \quad \text{and} \quad \ln \alpha_{it_p} = \beta_1 + \beta_2 GymAval_{it_p} + \beta_3 INC_{it},$$
where $INC_{it}$ is consumer $i$’s annual income in period $t$, $GymAval_{it}$ is a measurement of fitness equipments availability during period $t$, and $GymAval_{itp}$ is the availability during peak period $tp$. If we observe $GymAval_{it}$ and $GymAval_{itp}$, but do not observe $INC_{it}$, $X_{1it} = GymAval_{it}$ and $U_{it} = \beta_3 INC_{it} = U_{itp}$. On the other hand, if we observe $INC_{it}$, but do not observe $GymAval_{it}$ or $GymAval_{itp}$, $X_{1it} = INC_{it}$, $U_{it} = \beta_2 GymAval_{it}$, and $U_{itp} = \beta_2 GymAval_{itp}$. It is then unreasonable to claim that $U_{it} = U_{itp}$ because fitness gears availability is lower than usual.

5.2 Evaluate the money-metric effect of product changes

We have shown how to identify the demand function $D(p)$ in ??18. Demand function can be used to estimate the money-metric effect of product change, which will be useful given that product improvement is usually costly. Suppose a music streaming company wants to evaluate the effects of streaming music at high quality (HQ) which is more costly. For simplicity, suppose $X_{it} = (X_{1it}, X_{2it})'$, and $X_{1it}$ is a dummy variable that equals 1 if the music was streamed at HQ and 0 otherwise, and $X_{2it}$ is an excluded variable with large support. The company can run experiments in two cities with $X_{1it} = 0$ (non-HQ) in one city and $X_{1it} = 1$ (HQ) in the other. Our method can identify $F_W(W_{it} | X_{1it} = 1)$ and $F_W(W_{it} | X_{1it} = 0)$.

Then we have demand functions for the music streaming service with and without HQ, $D(P | X_{1it} = 1) = 1 - F_W(P | X_{1it} = 1)$ and $D(P | X_{1it} = 0) = 1 - F_W(P | X_{1it} = 0)$. Then $(D(P | X_{1it} = 1) - D(P | X_{1it} = 0)) / D(P | X_{1it} = 0)$ is the percentage change in the revenue when the music is streamed at HQ.

5.3 Pricing based on usage

Using individual usage data, we have shown that we can identify $F_W(W_{it} | X_{1it}, Q^*_{it})$, hence $F_W(W_{it} | Q^*_{it})$. Knowing consumer’s WTP given their usage, we can specify a pricing model based on usage and derive the corresponding demand curve and marginal revenue.

For simplicity, assume that there is no $X_{1it}$, the subscription choice follows $S_{it} = I(ln Q_{it}^* - ln P + U_{it} > 0)$, and $Q_{it}^* \perp U_{it}$. We have identified $F_U(u)$. Consider the following counterfactual usage based pricing model. At price $P_1$, one can use the subscription service up to $Q^h$ amount. With a higher price $P_2$, one has unlimited usage of the service. Assume that the distribution of the usage is unchanged in this counterfactual pricing model.\footnote{We first identify $F_W(w | X_{1it}, X_{2it})$, then we use $F_W(w | X_{1it}) = E(F_W(w | X_{1it}, X_{2it}) | X_{1it})$.}

\footnote{It can be problematical to make the assumption that the distribution of consumption does not change when consumers face the above counterfactual usage based pricing model. Nevo et al. (2016) have found consumers will adjust their consumption to adapt the threshold imposed by the three-part tariff pricing.}
Under this pricing model, the demand consists of the demand for the plan with quota $Q^h$ and price $P_1$ (called plan 1) and the plan without limit and price $P_2$ (called plan 2). The demand for plan 1 is proportional to

$$\int_{-\infty}^{\ln Q^h} P_A(q) f_{\ln Q^*}(q) \, dq + P_B \int_{\ln Q^h}^{\ln Q^h + \ln(P_2/P_1)} f_{\ln Q^*}(q) \, dq,$$

where

$$P_A(q) \equiv \Pr(q - \ln P_1 + U_{it} > 0) = 1 - F_U(\ln P_1 - q),$$

$$P_B \equiv \Pr(\ln Q^h - \ln P_1 + U_{it} > 0) = 1 - F_U(\ln P_1 - \ln Q^h).$$

Here $P_A(q)$ for $q < \ln Q^h$ is the probability that one would subscribe plan 1 given her log of expected usage is $q$, $P_B$ is the probability that one would subscribe plan 1 given that she would use the entire quota $Q^h$, and $\int_{\ln Q^h}^{\ln Q^h + \ln(P_2/P_1)} f_{\ln Q^*}(q) \, dq$ is the proportion of people in the population whose expected usage exceeds the quota $Q^h$ but find plan 2 too expensive hence will only consider plan 1. The demand for plan 2 is proportional to

$$\int_{\ln Q^h + \ln(P_2/P_1)}^{\infty} P_C(q) f_{\ln Q^*}(q) \, dq,$$

where

$$P_C(q) \equiv \Pr(q - \ln P_2 + U_{it} > 0) = 1 - F_U(\ln P_2 - q).$$

Here $P_C(q)$ is the probability that one would subscribe plan 2 given her log of expected usage is $q$. Knowing the demand function, it is straightforward to derive the marginal revenue and find the optimal $(P_1, P_2, Q^h)$ when the marginal cost is known.

6 What usage tracking data cannot tell us but price variation can

All of our previous results rely on the multiplicative form of the WTP $W_{it} = \alpha_{it} Q^*_{it}$. One natural question is that can we relax this multiplicative form? The answer of this question depends on whether or not we have price variation at all in data. First, we are going to argue that without any price variation, subscription/churn choices and usage tracking data cannot tell us the functional form of WTP $W_{it}$ as a function of usage $Q^*_{it}$. Moreover, without knowing the functional form, the distribution of the WTP cannot be identified. Secondly, we show how to use price variation to parametrically identify the functional form of $W_{it}$ as a function $Q^*_{it}$. We end this section with the identification of the outside option valuation as an application of the general result.
6.1 What usage tracking data cannot tell us

We focus on the case where we observe the expected usage for all sampled individuals. Suppose one lets
\[ W_{it} = g(\ln \alpha_{it} + \ln Q_{it}^*) = g(\beta'X_{1it} + U_{it} + \ln Q_{it}^*), \]
where \( g \) is some strictly increasing differentiable function. The multiplicative form \( W_{it} = \alpha_{it}Q_{it}^* = \exp(\ln \alpha_{it} + \ln Q_{it}^*) \) assumes that \( g \) is an exponential function. As a motivation for general functional form \( g \), consider the Taylor expansion of \( g(\ln \alpha_{it} + \ln Q_{it}^*) \) at \( \ln Q_{it}^* = 0 \). We have
\[ W_{it} = g(\ln \alpha_{it}) + g'(\ln \alpha_{it}) \ln Q_{it}^* + g''(\ln \alpha_{it})/2(\ln Q_{it}^*)^2 + \ldots, \]
which gives us nonlinear partial effect of usage on WTP.

Below we show that without knowing \( g \), one cannot identify the distribution of the WTP \( W_{it} \). To see this, let’s ignore \( X_{1it} \) and \( Z_{it} \), and let \( \ln \alpha_{it} = U_{it} \). We have
\[ S_{it} = \mathbb{I}(W_{it} > P) = \mathbb{I}(g(U_{it} + \ln Q_{it}^*) > P) = \mathbb{I}(U_{it} + \ln Q_{it}^* > g^{-1}(P)). \]
By the similar arguments in the proof of ??, we obtain the CDF of \( U_{it} \) as follows,
\[ F_U(u; g) = 1 - \mathbb{E}(S_{it} \mid \ln Q_{it}^* = g^{-1}(P) - u), \]
and the conditional CDF of the WTP \( W_{it} \) given \( Q_{it}^* \) as follows,
\[ F_W(w \mid Q_{it}^* = q; g) = 1 - \mathbb{E}(S_{it} \mid \ln Q_{it}^* = g^{-1}(P) - g^{-1}(w) + \ln q). \]
So the identified \( F_U(U_{it} ; g) \) and \( F_W(W_{it} \mid Q_{it}^* ; g) \) will change with respect to \( g \).

6.2 Benefits of price variation across markets

We have seen that certain tight specification, like multiplicative form in ??, 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. There could also be temporal price variation for the same market. For example, PlayStation Plus in the UK was priced at £39.99 before September 2017.

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 \( m \). Assume that the log of WTP for the service in market \( m = 1, \ldots, M \) equals,

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

Let \( g(\cdot; \delta) \) be a strictly increasing function known up to a finitely dimensional vector of parameters \( \delta \). 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

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

(6)

Note that \( \delta \) can include the valuation of the outside option, which could even be market specific when there is enough price variation within a market. As we will show in the next subsection that when the multiplicative form of WTP in \( \beta \) holds, we only need two distinct prices in one market to identify the valuation of the outside option in that market. So if the multiplicative form assumption holds and each market has more than two different prices, we can identify the valuation of the outside option for each market. For simplicity, assume \( X_{1it,m} \) is exogenous, so we ignore \( Z_{it} \) below. By the same arguments in the previous sections, we have

\[
\beta = E(X_{1it,m}X_{1it,m}')^{-1} E(X_{1it,m}Y_{2it,m}),
\]

where

\[
Y_{2it,m} = \frac{S_{it,m} I(\tilde{Q}_{it,m} - g^{-1}(P_m; \delta) > 0)}{f_{\tilde{Q}_m}(\tilde{Q}_{it,m} | X_{1it,m})}.
\]

For simplicity, suppose \( X_{1it,m} \) has the same distribution across different markets, hence \( E(X_{1it,m}X_{1it,m}') \) is identical across \( m \). We then have a set of moment equations of \( \delta \),

\[
E(X_{1it,j}Y_{2it,j}) - E(X_{1it,k}Y_{2it,k}) = 0, \quad 1 \leq j < k \leq M,
\]

(7)

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 \( \delta \) of \( g(\cdot; \delta) \). Once \( g \) is identified, the identification of the distribution of \( W_{it} \) follows from the earlier results.

### 6.3 Identify the utility of the outside option by price variation

As one application of the above arguments about the role of price variation, we show how to identify the WTP for the outside option, \( \mu \), by price variation within the same market. In some applications, we can observe multiple prices for the same subscription service for
various reasons. For example, in our music streaming empirical application, there are three common prices 99, 129 and 149 New Taiwan dollars (NT$) per month, among which 149 is the usual price, 129 is the price for customers of certain cellular carriers, and new users have special price 99 for the first two months.

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 following rule,

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

Here the outside option valuation \( \mu \) is unknown, but invariant of the prices. This specification is a special case of eq. (6) by letting \( g(c; \delta) = e^c - \delta \) and \( \delta = \mu \). We then can use the moment eq. (7) to identify \( \mu \). To be concrete, suppose \( X_{1it,m} = 1 \) only. Equation (7) reads

\[
E\left( \frac{S_{it,j}}{f_{\tilde{Q}_j}(\tilde{Q}_{it,j})} - \frac{S_{it,k}}{f_{\tilde{Q}_k}(\tilde{Q}_{it,k})} \right) = E\left( \frac{\mathbb{1}(\tilde{Q}_{it,j} - \ln(P_j + \mu) > 0)}{f_{\tilde{Q}_j}(\tilde{Q}_{it,j})} - \frac{\mathbb{1}(\tilde{Q}_{it,k} - \ln(P_k + \mu) > 0)}{f_{\tilde{Q}_k}(\tilde{Q}_{it,k})} \right),
\]

for any \( 1 \leq j < k \leq M \).

To see more intuition, it can be shown that

\[
\text{RHS of eq. (8)} = \ln \left( \frac{P_k + \mu}{P_j + \mu} \right),
\]

and

\[
\text{LHS of eq. (8)} = \int_{\eta} \int_{\eta} (S_{it,j} - S_{it,k}) \ dF_\eta(\eta) \ dq.
\]

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

7 Simulation Results

8 Empirical Application
References


