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

Cheng Chou Vineet Kumar*

November 2021

Abstract

We demonstrate how to obtain the distribution of consumer willingness to pay (WTP) for subscription products, where consumers pay a fixed price each period for potentially unlimited usage, e.g. Spotify. In the absence of price variation, we demonstrate how the variation in usage and subscription choice together can identify the WTP distribution. We apply our method to an empirical application using the data from a music streaming service. Using the estimated WTP distribution, we obtain the revenue maximizing prices for different consumer segments.

1 Introduction

Our paper studies how to obtain the distribution of consumer 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

*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 Associate Professor of Marketing at School of Management, Yale University. e-mail: vineet.kumar@yale.edu. 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.

value various features of the product, and to decide how alternative products should be priced. Consider the example of Spotify, which has a monthly Standard plan priced at $9.99 in the US. When the firm is interested in evaluating how demand might vary with price increases (i.e. price elasticity), we would need to obtain the WTP distribution so that we can infer the percentage of consumers who are still willing to pay more than the new higher 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 et al., 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 A1. 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).

In most subscription markets, prices are typically fairly stable (except for promotions like free trials). Spotify has always set the monthly price for unlimited ad-free streaming at $9.99 from 2011 to the present. Apple (Music and iCloud) and Microsoft (Office 365) are similar in terms of lack of price variation. While we might expect that digital technology reduces menu costs and makes firms more likely to change prices (Stamatopoulos, Bassamboo and Moreno, 2021), subscriptions firms are often especially wary of experimentation especially on price (Ariely, 2010). The reasons cited include wanting to avoid consumer confusion, consumer strategic timing or perceptions of unfairness among others. On the other hand, we often have access to high-frequency data about the usage of a subscription product (e.g. the amount of time spent in listening to Spotify at daily or hourly frequencies).

One of the crucially important decisions is pricing, which depends on the distribution of consumer WTP. Almost all extant research deals with obtaining WTP when prices vary, making it important to understand how WTP can be obtained in subscription markets. We examine the following research questions with subscriptions of digital entertainment services (e.g. streaming TV and music) as the empirical context. First, in an empirical setting without price variation, what can we infer about the distribution of consumer valuation of the product from usage and subscription data? Second, is usage variation equivalent to price variation in obtaining all economic primitives?

If not, what further inference is possible when we have price variation in addition to usage variation?

The essential feature of obtaining WTP from data (both observational and conjoint-like approaches) is that prices vary exogenously. This variation informs us of the shape of the demand curve. Demand estimation in economics and marketing has depended on the presence of data with price variation. Thus, 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? This 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, consumers have sometimes been found to have a different WTP when making actual purchase choices. Moreover, this approach does not get around the requirement for price variation. To the best of our knowledge, no existing research demonstrates the identification of WTP distribution without price variation.

The research contribution lies in our insight that purchase is separate from usage for subscription products—two Spotify subscribers paying the same price can have substantially different amount of usage. Because the price paid becomes a sunk cost at the beginning of the subscription period, a consumer chooses an optimal usage level according to her/her usage preference and available leisure time. When two subscribers pay the same monthly fee but have different usage level, these two consumers are paying different price per unit of usage, which opens up the opportunity of identifying the WTP distribution. We prove that the combination of usage and subscription data can identify the WTP distribution under a broad set of conditions. Overall, we propose a novel method to identify and estimate the conditional distribution of WTP given product features and customer characteristics when price variation is absent. We examine the other question of whether usage variation is a replacement for price variation, and find that while usage variation is helpful, it does not serve as a replacement for price variation in general.

Our framework is built upon a microfoundation-based model of product usage that occurs at high frequency.[1] The approach does not assume a specific functional form of the utility from using the subscribed product, but can accommodate common utility functions like Cobb-Douglas and constant elasticity of substitution (CES). The microfoundations involve a consumer trading off a latent leisure budget between product use and other options in each time period (day). To accommodate zero usage, the model allows for the possibility that consumers might not be able to use the service effectively unless they have a sufficient leisure budget available. The leisure budget is impacted by shifters, which provides exogenous variation and consequently shifts usage. Next, we connect the high-frequency (daily) usage to lower-frequency (monthly) purchase. We aggregate the expected indirect utility (conditional on the consumer's information set) from usage across the days to obtain a monthly expected utility. This monthly aggregate utility reflects daily utility maximization. The expected optimal utility for the month obtained from a subscription is then traded off against the price when the consumer makes a purchase decision. High-frequency shocks impact the daily usage decision, whereas low-frequency shocks impact the monthly subscription decision. The consumer in our model does not need to know future high-frequency (daily) shocks when making (monthly) purchase decisions, but knows the distribution of these shocks.

The identification involves two steps. First, we use observed daily usage to recover unobserved expected monthly leisure which is an aggregate of expected daily leisure. This is possible we prove that the daily usage equals the product of daily leisure and the share of leisure budget allocated to using the subscription. Modeling the share as varying across subscription periods (months) but fixed within them, we demonstrate how to recover the expected daily leisure from usage. Second, given the expected monthly leisure, we then use the variation of subscription choices to determine the distribution of consumer valuation of leisure. Intuitively, if two consumers had the same expected amount of leisure, and one subscribed but the other did not, it must be due to the difference between their valuations for leisure.

We provide a detailed estimation algorithm comprised of simple steps that uses commonly available data from subscription services to obtain the conditional WTP distribution. We first use the high-frequency usage data to estimate the parameters

[1] This is typically at daily data frequency but can often be measured with finer granularity.

relating usage to daily leisure shifters. In the parametric specification, we use a Tobit model for this step. We then estimate the leisure at the consumer-month level, which plays an important role in determining the subscription decision. This monthly leisure is not just an aggregation of daily leisure levels due to the possibility of zero usage. Next, we estimate the binary subscription decision related to monthly leisure and consumer and product characteristics using a probit model in the parametric specification. Having obtained the parameters of the usage model and subscription model, we can estimate the conditional WTP. The demand curve and other primitives like elasticity are obtained from the conditional WTP, and counterfactuals can then be performed.

Lastly, we take our method to data using an application of music streaming, featuring monthly subscription choices and daily usage (daily hours listening to streaming music) data. We estimate the distribution of WTP and price elasticities of the WTP for its current monthly streaming plan for different age and gender groups. 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 female subscribers are less price sensitive than male subscribers. Finally, using our estimates and model, we obtain the revenue maximizing prices for different consumer segments.

Our method is quite general, both in terms of the usage microfoundations and in terms of the overall purchase model. Although we examine the case of parametric identification in the main paper, we show non-parametric identification and estimation in Appendix B. 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 the method can be applied to typical packaged goods markets for instance.

The rest of the paper is organized as follows. Section 2 reviews the literature. In Section 3, we model a consumer's choices of whether or not to subscribe a product/service and the amount of usage of the subscription if subscribed. After the model setup, we discuss in Section 4 how to identify and estimate the model and to obtain the distribution of consumer WTP by leveraging the data of usage and subscription choices. We leave the extensions (the value of price variation, the effect of switching cost, and the effect of the entry of new service providers) to Section 5. Section 6 uses our approach in an application of music streaming subscription to demonstrate its empirical value. Section 7 concludes the paper. The appendix contains additional results about the nonparametric identification and estimation of the WTP distribution. The online appendix contains the technical proofs and a simulation study that examines the finite sample properties of the estimator.

2 Literature

There are multiple streams of literature focused on measuring and characterizing WTP or the distribution of consumer valuations. 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 hypothetical WTP. Such approaches are typically used to obtain hypothetical WTP since consumers do not have to actually pay a price or face financial consequences. Within this stream there are two broad approaches: direct surveys (Mitchell and Carson, 2013; Hanemann, 1994) and choice-based conjoint analysis (Green and Rao, 1971; Green and Srinivasan, 1978; Rao, 2014; Ding, 2007). 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. 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; 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 (e.g. Guadagni and Little, 1983), or market-level like in (e.g. Berry, 1994; Berry, Levinsohn and Pakes, 1995) and related literature. It is striking that *none of the above methods* provide any help when there is no price variation in the data. 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 even require multiple products. However, we do need access to usage data, consistent with our focus on subscription markets.

The closest paper we could find 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.

3 Model

3.1 Setup and Overview of the Main Results

We focus on the subscription of digital entertainment (e.g. streaming music and TV) as the empirical context. For concreteness, consider a monthly music streaming service. Let $i = 1, \dots, n$ index a consumer, and let $m = 1, \dots, M$ index a month. The sample has M months consisting of T days in total indexed by $t = 1, \dots, T$. Denote $m(t)$ the month containing day t. We observe consumer monthly binary subscription choice $S_{im} = 1$ (subscribing in month m) or 0 (not) and daily usage $Q_{it} \geq 0$ of the service if subscribed. In addition, we may observe consumer characteristics X_{im}.[2] Daily usage Q_{it} can be understood as the amount of time a consumer spends in listening to music using the subscription on day t. The data follow a cohort of consumers who were subscribing the service in the first sample month. So for all sampled consumers, we

[2]Let $X_{im} = 1$ if we do not observe any. We can also let X_{im} include observable product characteristics if available.

observe their usage for at least one month.

Consumer i makes a choice on whether or not to subscribe in month m at the beginning of the month by comparing the expected indirect monthly utility with a subscription W_{im} and the monthly subscription cost P:

$$S_{im} = \mathbf{1}(W_{im} - P > 0). \tag{1}$$

So W_{im} can be interpreted as the WTP or reservation price in month m, and $(W_{im} - P)$ is the consumer's surplus. In subscription settings, it is important to allow flexibility for WTP to vary month-to-month, due to the change of product (e.g. the release of new contents) or individual situation (e.g. student users have less leisure near the end of the semester).

The question is how to identify and estimate the distribution function $F_W(w)$ of W_{im} and other distributional features of W_{im}, such as its median? When price P does not change, the subscription choice S_{im} alone cannot identify the entire function $F_W(w)$; we only know the proportion of consumers who have WTP greater than the price, i.e. we know $F_W(P) = 1 - \Pr(S_{im} = 1)$ at the fixed price P by eq. (1).

What determines the distribution of the WTP for the service? Intuitively, the WTP for a music streaming service may vary across consumers because some consumers have more leisure hence they expect to use the service more and/or some consumers have a higher valuation of leisure activities therefore they are willing to pay more. Moreover, even for the same consumer, his or her WTP may vary over time due to product changes (e.g. the release of new contents) which affect the valuation of leisure activities including using the subscription. We model these parsimoniously by allowing for time-varying WTP. The primitives include a time-varying leisure process and a utility function of leisure activities.

We begin with an overview of the method. Our solution relies on the following expression of consumer WTP for the service in month m, i.e. W_{im}:

$$W_{im} = \alpha_{im} L_{im} \qquad \text{or equivalently} \qquad \ln W_{im} = \ln \alpha_{im} + \ln L_{im}, \tag{2}$$

where L_{im} is the expected amount of leisure in month m, and α_{im} is a parameter representing consumer i's valuation of leisure activities when she has a subscription.[3]

[3]More precisely, α_{im} is the maximum money-metric utility consumer i could obtain from 1 unit of leisure time when she subscribes and optimally allocate that leisure between using the subscription product and doing other leisure activities.

At first glance, it might appear that the linear form of the WTP might seem restrictive, we show that this form actually holds for a wide class of utility models that serve as a microfoundation of usage (see Theorem 1 below).

There are two sources of heterogeneity in consumer WTP in Equation (2): leisure amount (L_{im}) and the valuation of leisure activities including using the subscription (α_{im}). The two dimensional heterogeneity can accommodate two types of consumers of the music streaming service: subscribers who have more leisure hence expect more usage of the product like college students, and subscribers who are willing to pay more for listening to music though they may have lower usage due to less leisure time, e.g. professionals like lawyers. Broadly, if we assume that utility (and WTP) is higher across consumers with higher levels of usage, we would be conflating these two underlying factors resulting in biased estimates. In our empirical study of streaming music, we in fact find that though the older consumers use less, they are indeed willing to pay more for the subscription.

By decomposing WTP W_{im} into two components (α_{im} and L_{im}), we can combine the information from both subscription choices (for α_{im}) and usage (for L_{im}). First, we will prove a formula for observed usage in terms of unobserved leisure (in part (2) of Theorem 1 below). This formula is crucial in recovering the expected leisure L_{im}. This step involves only usage data. Second, knowing the expected monthly leisure L_{im}, we only need the distribution of α_{im} in order to find the distribution of $W_{im} = \alpha_{im} L_{im}$, provided that α_{im} and L_{im} are independent (denoted by $\alpha_{im} \perp\!\!\!\perp L_{im}$). To be clear, our method *does not* require this independence assumption, but we use it only in this overview to make the logic and intuition transparent.

This second step uses data from subscription choices. To see how, note that eq. (1) can be written as $S_{im} = \mathbf{1}(\alpha_{im} > P/L_{im})$ when $W_{im} = \alpha_{im} L_{im}$. If $\alpha_{im} \perp\!\!\!\perp L_{im}$, we have $\Pr(\alpha_{im} \leq a) = \mathrm{E}(1 - S_{im} \,|\, L_{im} = P/a)$ for any value a,[4] and the conditional expectation is known because S_{im} is observed, and L_{im} can be recovered from usage. Below, we will add the modeling details of how we address the correlation between α_{im} and L_{im} and how to incorporate observed consumer heterogeneity X_{im}. The key condition is that there exists some exogenous variables that will change expected leisure L_{im} but not preference α_{im}.

[4]By $S_{im} = \mathbf{1}(\alpha_{im} > P/L_{im})$, we have $\Pr(1 - S_{im} = 1 \mid L_{im} = P/a) = \Pr(\alpha_{im} \leq P/(P/a) \mid L_{im} = P/a) = \Pr(\alpha_{im} \leq a)$. The last identity used the condition $\alpha_{im} \perp\!\!\!\perp L_{im}$.

3.2 Microfoundations of Usage

The monthly utility of a subscription is built up from the indirect utility that is obtained from the daily usage of the product. We adopt a money-metric representation of the daily direct utility a consumer receives from her leisure time spent in listening to streaming music, denoted by q_{it}, and in doing other leisure activities (e.g. watching TV), denoted by q_{0it}.[5]

For any day t, consumer i chooses $(q_{it}, q_{0it})'$ to maximize her utility from leisure activities:

$$\max u_{im(t)}(q_{it}, q_{0,it}) \qquad \text{subject to} \qquad q_{it} + q_{0,it} = \ell_{it}, \tag{3}$$

where $\ell_{it} \geq 0$ denotes the unobservable (to researchers) leisure time on day t. Recall $m(t)$ is the month containing day t. By writing $u_{im(t)}$, we allow the utility function to change across consumers and months.

We model the daily leisure $\ell_{it} \geq 0$ using a Tobit specification in order to allow the following two features. First, the leisure generated from this equation is non-negative. Second, the model needs to accommodate the possibility of observing zero daily usage, which is common in data. In the Tobit model, there is a positive probability that the leisure is zero, which corresponds to zero usage of service. Intuitively, this says that some days the consumer has some outside activities (working, childcare, etc.) that does not permit them the luxury of using the focal service.

The daily leisure $\ell_{it} \geq 0$ is specified as follows:

$$\ell_{it} = \begin{cases} \ell_{it}^* & \text{if } \ell_{it}^* > 0 \\ 0 & \text{if } \ell_{it}^* \leq 0, \end{cases} \tag{4}$$

where ℓ_{it}^* is a latent variable, and

$$\ell_{it}^* = \gamma' Z_{it} + \mu_i + \varepsilon_{it}.$$

Here Z_{it} denotes a vector of exogenous covariates that affects leisure (e.g. weekend or holiday dummy variables or weather). These variables ultimately affect the usage of

[5] Money-metric utility functions are commonly used in the study of the WTP for non-market goods or service, such as the amenities of school and neighborhood (Altonji and Mansfield, 2018); money-metric utility functions also have a long history in the literature of hedonic models (starting from the seminal paper by Rosen, 1974), which serve as the workhorse model in estimating the WTP for amenities (e.g. neighborhood racial composition, violent crime, and air pollution) in housing market (Bayer et al., 2016) and the WTP for product features (Bajari and Benkard, 2005).

subscription. Note that μ_i is the unobserved consumer heterogeneity in the amount of leisure (e.g. age, gender, household size).

We normalize the utility of not using the service to be 0. Consumers do not have perfect foresight; particularly, they do not know exactly their amount of leisure in future days but rather form expectations over the leisure. When making a subscription decision for a month, the consumer must form expectations of leisure for all days in the month. The following assumption specifies the information available to consumers at the beginning of each month m, conditional on which they infer the expected amount of leisure in month m.

Assumption 1 (Consumer's Belief). *(1) At the beginning of month m, consumer i knows $\mathbf{Z}_{im} \equiv \{Z_{it} : m(t) = m\}$, the leisure parameters μ_i and γ, and her utility function u_{im} in month m.*

(2) The leisure shocks ε_{it} are independent and identically distributed across individuals and days with a normal distribution. Normalize $\text{Var}(\varepsilon_{it}) = 1$.[6] The leisure shocks ε_{it} are independent of $\mathbf{Z}_{im}$.

This basic assumption characterizes consumer knowledge at the time they make a subscription purchase. Consumers form forecasts of exogenous variable evolution over the period of the subscription. The leisure shocks serve to rationalize usage patterns, and the above assumption indicates that consumers cannot predict future leisure shocks. The above assumption allows us to characterize the optimal usage at the daily level and the corresponding indirect utility at the monthly level for a wide class of daily usage utility functions.

Theorem 1. *Suppose the observed daily usage Q_{it} is derived from microfoundations in Equation* (3) *trading off between using the subscription and doing other leisure activities, the monthly utility is additively separable in daily utility, and Assumption 1 holds. Then, for the class of (daily) utility functions that are homogeneous of degree 1 (including Cobb-Douglas, CES, perfect substitutes, perfect complements, Leontief, etc.), we have the following results:*

[6] We will argue later that this normalization is innocuous for the identification of the distribution of WTP.

(1) The expected monthly indirect utility W_{im} satisfies

$$W_{im} = \alpha_{im} L_{im} \qquad \textit{or equivalently} \qquad \ln W_{im} = \ln \alpha_{im} + \ln L_{im},$$

where L_{im} is the expected monthly leisure. L_{im} is a function of the conditioning variables $\mathbf{Z}_{im}$ and unobserved heterogeneity in leisure amount μ_i, and it equals the following,

$$L_{im} \equiv \sum_{t:m(t)=m} \left[\left(\mu_i + \gamma' Z_{it} + \frac{\phi(\mu_i + \gamma' Z_{it})}{\Phi(\mu_i + \gamma' Z_{it})} \right) \Phi(\mu_i + \gamma' Z_{it}) \right]. \tag{5}$$

Here ϕ and Φ denote the probability density function (PDF) and the cumulative distribution function (CDF) of the standard normal distribution.[7]

(2) The daily usage of the subscription satisfies

$$Q_{it} = r_{im(t)} \ell_{it}$$

for a preference parameter $r_{im(t)}$. The interpretation of $r_{im(t)}$ is the share of leisure budget spent in listening to streaming music. Neither $\alpha_{im(t)}$ nor $r_{im(t)}$ involves the leisure budget ℓ_{it}.

We note that the critical assumption required for our method, the linear relationship between WTP W_{im} and the monthly expected leisure L_{im}, holds for a class of common utility functions. Most empirical research in economics and marketing use these utility functions in their modeling of usage, reassuring us of the generality of scope. The interpretation of α_{im} is the maximum daily utility one could obtain from 1 unit of leisure time by optimally allocating the leisure time to the two activities. We also allow for flexibly modeling usage utility to be time-varying at a monthly level (through the $r_{im(t)}$ parameter) to rationalize that consumers might have seasonal variations in usage. When usage data is available at a higher frequency than purchase data, this becomes especially useful in capturing such temporal variations.

We use the familiar Cobb-Douglas utility function to illustrate the general conclusion in the above theorem, and point out why the preference parameter α_{im} could be correlated with the expected monthly leisure L_{im}.

[7] This formula comes from the Tobit specification of the daily leisure. In particular, $\Phi(\mu_i + \gamma' Z_{it})$ is the probability that $\ell_{it} > 0$, and $\phi(\mu_i + \gamma' Z_{it})/\Phi(\mu_i + \gamma' Z_{it})$ is the inverse Mills ratio.

Example 1 (Cobb-Douglas Utility). *Consider a Cobb-Douglas utility function*

$$u_{im(t)}(q_{it}, q_{0,it}) = \eta_i \cdot \left(q_{it}^{r_{im(t)}} q_{0,it}^{1-r_{im(t)}} \right).$$

The coefficient $r_{im(t)}$ is the marginal rate of substitution (MRS) between the two leisure activities (listening to streaming music and watching TV), which depends on individual preference and product characteristics. Because the product characteristics (e.g. the number of shows) could change over time, we let the MRS to be time varying. When we adopt a money-metric representation of the utility from leisure activities, it is natural to incorporate the possibility that people would assign different dollar values to their utilities of leisure due to heterogeneity such as wage rates. The parameter η_i, which can be viewed as a function of wage rate according to the neo-classical economics theory (e.g. chapter 4 of Deaton and Muellbauer, 1980), is to capture such heterogeneous valuation of leisure—the η_i of professional lawyers is higher than the η_i associated with students.

We have that the optimal amount time of listening to streaming music is

$$Q_{it} = r_{im(t)} \ell_{it},$$

which is the second conclusion of Theorem 1. The optimal amount time of watching TV is then $Q_{0,it} = (1 - r_{im(t)})\ell_{it}$. Particularly, for one unit of leisure, $r_{im(t)}$ and $1 - r_{im(t)}$ are the optimal amount of time spent in music and TV, respectively. The indirect utility on day t can be expressed in terms of ℓ_{it} as follows,

$$V_{it} \equiv u_{im(t)}(Q_{it}, Q_{0,it}) = \overbrace{\left[\eta_i {r_{im(t)}}^{r_{im(t)}} \left(1 - r_{im(t)}\right)^{1-r_{im(t)}}\right]}^{\equiv \alpha_{im(t)}} \ell_{it}.$$

Let the term in the bracket be $\alpha_{im(t)}$ in Theorem 1. So the interpretation of $\alpha_{im(t)}$ is the money-metric value of one unit of leisure when the consumer has a subscription and is optimally trading off between using the subscription and doing other activities; it depends both a consumer's valuation of leisure and preference regarding alternative leisure activities.

Because the monthly utility is additively separable in the daily utility, the monthly utility is the sum $\alpha_{im} \sum_{t:m(t)=m} \ell_{it}$. In order to obtain the expected *monthly utility, we need to take the conditional expectation of the total leisure $\sum_{t:m(t)=m} \ell_{it}$ given the information at the beginning of month m as described in Assumption 1. In the proof of Theorem 1, we show the expected total leisure has the form in eq.* (5).

In this Cobb-Douglas utility function, α_{im} depends on the MRS $r_{im(t)}$ between the two activities (listening to streaming music and watching TV) and the dollar value assigned to the utility from leisure (η_i). The latter (η_i) is presumably correlated with one's wage rate, which is further related to her expected leisure L_{im}. The MRS $r_{im(t)}$ can also be correlated with L_{im}. For example, the MRS is affected by whether or not the consumer has a Netflix subscription. The consumer decision about subscribing Netflix intuitively will also depend on her expected leisure L_{im}. So in general we expect that α_{im} and L_{im} are correlated.

Up to now, we have shown the decomposition $W_{im} = \alpha_{im} L_{im}$ or equivalently $\ln W_{im} = \ln \alpha_{im} + \ln L_{im}$.

Consumer Heterogeneity and Correlation: We now focus on two other aspects of the model that allows it to be more realistic. First, we show how to incorporate the observed consumer heterogeneity X_{im} into the indirect utility and consequently the purchase decision. This is important since the value of leisure α_{im} may depend on consumer characteristics, in addition to time-varying unobservables. Second, we show how the model can incorporate correlation between value of leisure and expected monthly leisure, L_{im}. This correlation is important, for instance, if we expect that in months that consumers have more leisure, they might have income shocks that also impact their value of leisure, and in turn, their WTP.

We first detail how we take account of observed consumer heterogeneity X_{im}. Consider a linear projection of $\ln \alpha_{im}$ onto X_{im} as:

$$\ln \alpha_{im} = \beta' X_{im} + U_{im} = \beta_0 + \beta_1' X_{1im} + U_{im}, \tag{6}$$

where $\beta' = (\beta_0, \beta_1')$ and $X_{im}' = (1, X_{1im}')$.[8]

The residual U_{im} can be interpreted as the unobserved consumer heterogeneity in the valuation of leisure activities with an active subscription after controlling for the observed factors X_{im} that could be both time-varying and heterogeneous. Because $\ln W_{im} = \ln L_{im} + \ln \alpha_{im}$, we have

$$\ln W_{im} = \ln L_{im} + \beta' X_{im} + U_{im}. \tag{7}$$

[8]By the *definition* (not an assumption) of linear projection (Wooldridge, 2010, pg. 25), $\beta_1 = [\mathrm{Var}(X_{1im})]^{-1} \mathrm{Cov}(X_{1im}, \ln \alpha_{im})$, and $\beta_0 = \mathrm{E}(\ln \alpha_{im}) - \mathrm{E}(X_{1im})'\beta_1$. The residual U_{im} has mean zero and is uncorrelated with X_{im}.

This equation says that β can be interpreted as the semi-elasticity of WTP with respect to the change of X_{im}, other things being same. Moreover, the binary subscription $S_{im} = \mathbf{1}(\ln W_{im} > \ln P)$ becomes

$$S_{im} = \mathbf{1}(\ln L_{im} + \beta' X_{im} - \ln P + U_{im} > 0). \tag{8}$$

This equation resembles the familiar threshold crossing binary choice model, though the log of expected monthly leisure is unobserved.

Consider the interpretation of β and U_{im} using the Cobb-Douglas utility function as an example.

Example 1 (continued)**.** *In the Cobb-Douglas utility function, we have seen that* $\alpha_{im} = \eta_i r_{im}^{r_{im}}(1 - r_{im})^{1-r_{im}}$*, and* η_i *depends on one's wage rate. For simplicity, suppose the MRS* r_{im} *between listening to music and watching TV is a constant* r *across consumers and over time. We then have*

$$\ln \alpha_{im} = [r \ln r + (1 - r) \ln(1 - r)] + \ln \eta_i.$$

If the data do not have any observed consumer heterogeneity, $X_{im} = 1$*, we have the following after the linear projection*

$$\ln \alpha_{im} = \overbrace{[r \ln r + (1 - r) \ln(1 - r) + \mathrm{E}(\ln \eta_i)]}^{\beta_0} + \overbrace{[\ln \eta_i - \mathrm{E}(\ln \eta_i)]}^{U_{im}}.$$

It is clear that β_0 *in this example is the population mean of the log money-metric value of one unit of leisure when one subscribes. Because* $\mathrm{E}(\ln \eta_i)$ *is the mean of (log) valuation of leisure,* U_{im} *is the individual deviation from the mean valuation of leisure.*

In the above example, we have seen that it is possible that α_{im} and L_{im} are correlated. It would be easier to assume that they are uncorrelated, but that would lead to inaccurate inference. The observed consumer heterogeneity X_{im} explains part of the correlation between α_{im} and L_{im}. When the correlation between α_{im} and L_{im} is due to the unobserved heterogeneity (such as unobserved wage rate), we have to rely on an exogenous shifter of leisure, Z_{it}.

Endogeneity: We detail the necessary exogenous variations required for identification in Assumption 2 below. This assumption allows for the correlation between leisure fixed effect μ_i and unobserved preference heterogeneity U_{im} across consumers for any given month m.

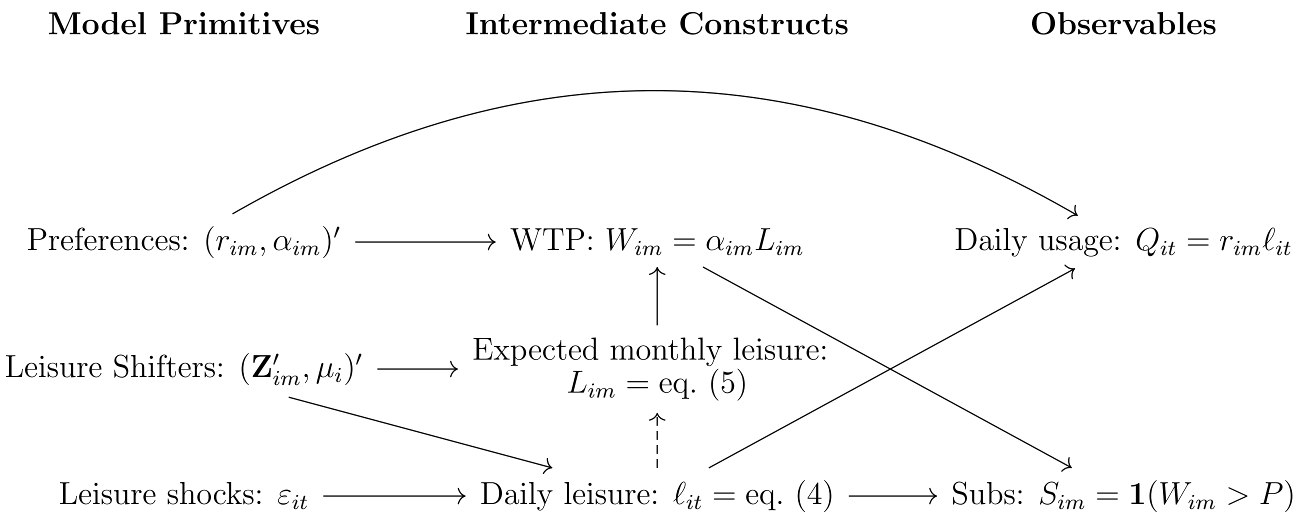

Figure 1: Conceptual Model Schematic

Assumption 2 (Exogenous Variation in Leisure)**.** *Assume that* $\mathbf{Z}_{im} \perp\!\!\!\perp U_{im} \,|\, (X_{im}, \mu_i)$*, which implies* $L_{im} \perp\!\!\!\perp U_{im} \mid (X_{im}, \mu_i)$ *because the randomness of* L_{im} *only comes from* $\mathbf{Z}_{im}$ *and* μ_i*.*

To understand why Assumption 2 is necessary, consider the case where L_{im} is known to us. According to linear expression of $\ln W_{im}$ in eq. (7), we need to know β in order to obtain the distribution of WTP W_{im}.

We typically have to use the binary subscription choice S_{im} in eq. (8) to obtain β. However, when the regressor L_{im} is correlated with U_{im}, we have the familiar endogenous regressor problem and any estimate of β obtained will be biased. To address this endogeneity issue, we typically obtain instrumental variables (IV) that affects leisure L_{im}, the endogenous regressor, but not the error term U_{im}, the unobserved preference heterogeneity. The instruments $\mathbf{Z}_{im}$ we suggest later in the application in Section 6 involves precisely this type of variable.

Summary of the Conceptual Model: We summarize the mechanism of our model with a schematic in Figure 1. The model primitives impact Intermediate Constructs, and both of these generate the observed data. From left to right of Figure 1, the model primitives consist of preference parameters $(r_{im}, \alpha_{im})'$, observed and unobserved leisure shifters $(\mathbf{Z}'_{im}, \mu_i)'$, and daily leisure shocks ε_{it}. The leisure shifters and daily leisure shocks determine the daily amount of leisure ℓ_{it}. Summing up the daily leisures for all days in one month and taking the expectation, we have the expected monthly leisure L_{im}, which is a function of the leisure shifters. The expected monthly leisure together with the preference parameters determines the WTP W_{im}. We observe daily usage of the subscription Q_{it} and binary monthly subscription choices S_{im}. The daily usage Q_{it} equals the daily leisure ℓ_{it} multiplied by the share of leisure budget spent on the subscription r_{im}. The subscription choice S_{im} is a result of comparing the WTP W_{im} and the subscription cost P.

4 Identification and Estimation of WTP Distribution

The objective is to identify and estimate the distribution function of W_{im} or equivalently its monotone transformation $\ln W_{im}$. We have seen that $\ln W_{im}$ has a linear additive form,

$$\ln W_{im} = \ln L_{im} + \beta' X_{im} + U_{im}.$$

We first discuss the identification strategy, which proceeds in two steps. In the first step, we use the observed daily usage $Q_{it} = r_{im}\ell_{it}$ to recover the parameters γ and μ_i inside the daily leisure ℓ_{it}. Knowing γ and μ_i, we know the expected monthly leisure L_{im} by its formula in eq. (5). In the second step, we identify β and the conditional distribution of U_{im} given (X_{im}, L_{im}) from the monthly subscription choices $S_{im} = \mathbf{1}(\ln W_{im} > \ln P)$. Then the distribution of $\ln W_{im}$ is recovered by the above linear additive form.

4.1 Identification

Step 1: Usage By the formula that $Q_{it} = r_{im(t)}\ell_{it}$ and the Tobit specification of the daily leisure process ℓ_{it} in eq. (4), the observed daily usage Q_{it} can also be written as a Tobit model:

$$Q_{it} = \begin{cases} Q_{it}^*, & Q_{it}^* > 0 \\ 0, & Q_{it}^* \leq 0, \end{cases} \tag{9}$$

where

$$Q_{it}^* = (r_{im(t)}\gamma)' Z_{it} + r_{im(t)}\mu_i + r_{im(t)}\varepsilon_{it}.$$

This is a parametric Tobit model with $\varepsilon_{it} \sim \mathcal{N}(0,1)$, from which we can identify (r_{im}, μ_i, γ') for each month m and consumer i. So $(r_{im(t)}, \mu_i, \gamma')$ is identified using *only* usage data, including the exogenous leisure shifter Z_{it}, but without requiring any subscription data. Consequently, the expected monthly leisure L_{im} is identified with only usage data.

Step 2: Subscription We next consider the identification of preference parameters β and the distribution of U_{im} from the subscription choice:

$$S_{im} = \mathbf{1}((\ln L_{im} - \ln P) + \beta' X_{im} + U_{im} > 0).$$

Note that after the first step, L_{im} is identified and can be viewed as known. Since the constant price P is known as well, it remains to identify β and the distribution of the unobservable U_{im}.

We focus on the parametric identification by assuming that the conditional distribution of U_{im} given (X_{im}, μ_i) is a normal distribution. With the normal distribution assumption, the binary choice of S_{im} is the standard probit model, from which we can identify the unknown parameters (see Theorem 2 below). We demonstrate that the distribution of the WTP and β are *nonparametrically identified*, i.e. the joint distribution of (X_{im}, μ_i, U_{im}) can be left unrestricted for each month m (see Theorem B.1 in the appendix).[9] Given Theorem B.1, we can demonstrate that our source of identification comes from the exogenous variation of $\mathbf{Z}_{im}$ rather than imposing particular parametric assumptions. However, we focus on the parametric form below, because it is more likely to be used in applications and also conveys the essential intuition that is more generally applicable.

Assumption 3 (Normal Distribution)**.** *(1) For each month m, assume that*

$$U_{im} \mid (X_{im}, \mu_i) \sim \mathcal{N}(\sigma_{u,\mu}\mu^*_{im}, \sigma^2_u),$$

where μ^*_{im} is the residual of the linear projection of μ_i onto X_{1im}.[10]

(2) Let $R_{im} \equiv (X'_{im}, \ln L_{im}, \mu_i)'$. Assume that $\mathrm{E}(R_{im}R'_{im})$ *is of full rank.*

[9] We also provide a simple formula for $\mathrm{E}(W_{im} | X_{im}, \mu_i)$ and $\mathrm{E}(\ln W_{im} | X_{im}, \mu_i)$, and that elasticities β can be estimated by an OLS estimator.

[10] That is $\mu^*_{im} = \mu_i - \mathrm{E}(\mu_i) + \sigma_{\mu,x}\Omega^{-1}_{x1}(X_{1im} - \mathrm{E}(X_{1im}))$, where $\sigma_{\mu,x1} = \mathrm{Cov}(\mu_i, X_{1im})$, and Ω_{x1} is the covariance matrix of X_{im}.

This conditional normal assumption is widely used in the correlated random effect model (see Chamberlain, 1980). We assume that the conditional mean of U_{im} given (X_{im}, μ_i) depends on the *residual* of the linear projection of μ_i onto X_{1im}. This is because X_{1im} is uncorrelated with U_{im} by the construction of the linear projection of $\ln \alpha_{im}$ onto X_{im}; given X_{1im}, U_{im} will only be correlated with the part of μ_i that is uncorrelated with X_{1im} (i.e. μ^*_{im}). The estimate of μ^*_{im} is the residual after running the linear regression of μ_i on X_{im} for each month m using all consumers $i = 1, \ldots, n$.

Part (2) of Assumption 3 makes the role of $\mathbf{Z}_{im}$ in the parametric identification clear. When we do not have access to the instrumental variable $\mathbf{Z}_{im}$ and μ_i is large so that the latent leisure variable ℓ^*_{it} is greater than 0, $L_{im} \approx \mu_i T_m$ (T_m is the number of days in month m) and R_{im} becomes $(X'_{im}, \ln L_{im} = \ln \mu_i + \ln T_m, \mu_i)'$. Because $\ln \mu_i$ and μ_i are highly collinear, the rank condition is unlikely to be satisfied.

Under Assumption 1 to 3, we have

$$\Pr(S_{im} = 1 \mid X_{im}, \mu_i, L_{im}) = \Phi\left(\frac{1}{\sigma_u}\ln(L_{im}/P) + \frac{\beta'}{\sigma_u}X_{im} + \frac{\sigma_{u,\mu}}{\sigma_u}\mu^*_{im}\right).$$

We can view the binary subscription choice S_{im} as the binary outcome, and view $\ln(L_{im}/P)$, X_{im}, and μ^*_{im} as the explanatory variables. The usual probit regression identifies the parameters σ_u^{-1}, β/σ_u, $\sigma_{u,\mu}/\sigma_u$. Then β and $\sigma_{u,\mu}$ are obtained easily by transformation. This is our conclusion in part (1) of Theorem 2 below. Knowing the parameters $(\beta, \sigma_u, \sigma_{u,\mu})$, we know the conditional distribution of U_{im} given (X_{im}, μ_i) by Assumption 3. We then can derive the distribution of the WTP W_{im} easily by using $F_W(w \mid X_{im}, \mu_i, L_{im}) = \Pr(\ln W_{im} \le \ln w \mid X_{im}, \mu_i, L_{im})$ and that $\ln W_{im} \le \ln w$ is equivalent to $U_{im} \le \ln w - \ln L_{im} - \beta' X_{im}$ because $\ln W_{im} = \ln L_{im} + \beta' X_{im} + U_{im}$.

Theorem 2 (Parametric Identification of WTP)**.** *Suppose Assumption 1 to 3 hold. We have*

1. *The unknown parameters $(\beta, \sigma_u, \sigma_{u,\mu})$ are identified.*

2. *The distribution of WTP is identified, and*

$$F_W(w \mid X_{im}, \mu_i, L_{im}) = \Phi\left[\frac{1}{\sigma_u}\left(\ln w - \ln L_{im} - \beta' X_{im} - \sigma_{u,\mu}\mu^*_{im}\right)\right].$$

As one particular application of the above theorem, we detail the estimation of the price elasticity e_{price} without price variation:

$$e_{price} = -\frac{\partial F_W(P)}{\partial P}\frac{P}{1 - F_W(P)}.$$

Using the expression of $F_W(w \mid X_{im}, \mu_i, L_{im})$ in Theorem 2, we have that

$$\begin{aligned} e_{price} &= -\frac{1}{\sigma_u \Pr(S_{im})} \int \phi \left[\frac{1}{\sigma_u} \left(\ln P - \ln L_{im} - \beta' X_{im} - \sigma_{u,\mu}\mu^*_{im}\right)\right] \mathrm{d}\, F(X_{im}, \mu_i, L_{im}) \\ &\approx -\frac{1}{\sigma_u \Pr(S_{im})} \frac{1}{nM} \sum_{i=1}^{n} \sum_{m=1}^{M} \phi \left[\frac{1}{\sigma_u} \left(\ln P - \ln L_{im} - \beta' X_{im} - \sigma_{u,\mu}\mu^*_{im}\right)\right]. \end{aligned} \tag{10}$$

The approximation follows from using the sample analog to estimate the integral. Note that the above elasticity e_{price} is the "overall" price elasticity across all consumers and all months. One of the advantages of our approach is that we can obtain WTP for different segments. Because we have identified the *conditional expectation of WTP* $F_W(w \mid X_{im}, \mu_i, L_{im})$, it is straightforward to compute the price elasticity for different consumer segments (such as students subscribers) and different months (e.g. holidays). In the empirical analysis, we will demonstrate the managerial value of these elasticities by considering the pricing of the subscription for different consumer segments.

Importance of Usage Data: We have now shown the identification when we have both usage and subscription data. To better understand the results, it is helpful to consider the consequence when we do not observe usage. In the absence of usage data, we will be unable to obtain the parameters (μ_i, γ) in daily leisure and consequently the expected monthly leisure L_{im}. The binary subscription equation

$$\begin{aligned} S_{im} &= \mathbf{1}(\ln L_{im} - \ln P + \beta' X_{im} + U_{im} > 0) \\ &= \mathbf{1}[(\beta_0 - \ln P) + \beta_1' X_{1im} + (\ln L_{im} + U_{im}) > 0] \end{aligned}$$

now involves two unknown error terms $\ln L_{im}$ and U_{im}. In such a situation with only subscription data, even if we made stronger distributional assumption that the sum $(\ln L_{im} + U_{im})$ follows a normal distribution with unknown variance, we can at most identify β up to scale and cannot identify the variance of $(\ln L_{im} + U_{im})$, which is actually essential even for the simple task like inferring the mean of the WTP W_{im}, which follows a log-normal distribution. So it is not possible to obtain WTP without observing usage, highlighting the unique role played by usage data.

4.2 Estimation

The estimation procedure is developed from the two step identification arguments with one noteworthy difference. In estimating the Tobit model of usage, we use a finite mixture model by assuming that there are a finite number of latent types of $(r_{im(t)}, \mu_i)'$. The reason why we have to take the approach of finite latent types is the following. If we did not group consumers by their latent types, the estimation of individual $(r_{im}, \mu_i)'$ using the Tobit model of usage will have to rely only on the number of observed days with active subscription for consumer i. For a consumer who cancelled her subscription after the first month, we only have about 30 (days) observations. This limited number of observations leads to estimation error in the estimate of μ_i, that enters into the estimate of L_{im}. The challenge is that the estimated L_{im} (containing the nonignorable estimation error) acts as a regressor in the second step probit regression of S_{im} on $\ln(L_{im}/P)$, X_{im} and μ^*_{im}. Consequently, the nonignorable estimation error inside the regressor L_{im} works like the measurement error in the regressors of a regression. It is well known that the measurement error, even classic ones, will bias the estimates of regression coefficients. We could potentially retain only the consumers who remain subscribers for a longer period, but that would introduce selection issues. To avoid these issues, we use latent classes (or types). By using the latent types, we can pool the information from a large number of consumers that will make the estimation error ignorable. In marketing the use of latent class models for the purpose of segmentation in choice models has a long history beginning with Kamakura and Russell (1989).

In practice, we use the expectation-maximization (EM) algorithm to estimate the finite mixture Tobit model of usage. From the EM algorithm, we obtain the estimates of (γ, μ_i). We then compute L_{im}. The last step is to run probit model to obtain the rest of parameters.

We conclude this section with the following estimation algorithm.

1. Estimate the finite mixture Tobit model eq. (9) by the EM algorithm. Let $(\hat{\mu}_i, \hat{\gamma}')'$ be the estimates of $(\mu_i, \gamma')'$ after running the EM algorithm.

2. Estimate L_{im} for each consumer and month by substituting the unknown parameters $(\mu_i, \gamma')'$ with the estimates $(\hat{\mu}_i, \hat{\gamma}')'$. Denote this estimator by $\hat{L}_{im}$.

3. For each month m, implement a linear regression of $\hat{\mu}_i$ on X_{im} and save the

residuals $\hat{\mu}^*_{im}$. These residuals are the estimates of μ^*_{im}.

4. Run the probit regression of S_{im} on $\ln(\hat{L}_{im}/P)$, X_{im}, and $\hat{\mu}^*_{im}$. The probit regression provides estimates of σ_u^{-1}, β/σ_u, $\sigma_{u,\mu}/\sigma_u$. Then the estimates of β and $\sigma_{u,\mu}$ are obtained easily.

Given the sequential nature of the routine, we recommend using bootstrap to obtain the standard error. In the Online Appendix, we conduct a numerical study and demonstrate the finite sample performance of the estimation algorithm.

5 Where is Price Variation Useful?

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

$$S_{im} = \mathbf{1}(W_{im} > P_{im} + \delta' X_{2im}),$$

where X_{2im} is a vector of observable covariates, and P_{im} denotes the price faced by consumer i in month m. We can interpret $P_{im} + \delta' X_{2im}$ as the total cost of a monthly subscription (e.g. price and switching cost). We write price P_{im} to analyze the general case in which price may or may not vary. For simplicity of discussion, assume X_{1im} and X_{2im} are not overlapping, and let $X_{im} = (1, X'_{1im}, X'_{2im})'$ in this extension.[11] We have seen the special case $P_{im} = P$ and $\delta = 0$. We maintain our assumption (Assumption 1) about consumer's utility of using the subscribed service and leisure, so that the conclusion $W_{im} = \alpha_{im} L_{im}$ in Theorem 1 holds. Using $\ln W_{im} = \ln L_{im} + \beta' X_{im} + U_{im}$, we can write the subscription decision in this more general setting as:

$$S_{im} = \mathbf{1}(\ln L_{im} - \ln(P_{im} + \delta' X_{2im}) + \beta' X_{im} + U_{im} > 0).$$

[11] When X_{1im} and X_{2im} overlap, one can easily modify the proof of Theorem 3 below by (a) defining a new notation, say $\tilde{X}_{2im}$, for the vector of variables that appear in X_{2im} but not in X_{1im}, and (b) substituting the occurrence of X_{2im} in the proof with $\tilde{X}_{2im}$. We did not pursue this cumbersome exposition since our current arguments sufficiently achieve the main objective of clarifying the information of price variation.

To see the motivation of this general case, we provide two examples. These two examples are not only interesting by themselves but also showcase different scenarios of identification with and without price variation.

Case 1 (Entry of New Platform)**.** *Suppose our data are about the subscribers of Spotify. Apple launched Apple Music, its streaming music subscription, on June 30, 2015. It is helpful to understand how our model accounts for the entry of Apple Music, and how this entry decision impacts the demand for Spotify. If we have data that includes the months before and after the launch of Apple Music, we can create a dummy variable* $Apple_{im}$ *that equals* 1 *for the months after June, 2015 and* 0 *before. The entry of Apple Music changes the value of the outside option. So the subscription rule becomes*

$$\begin{aligned} S_{im} &= \mathbf{1}(W_{im} > P + \delta Apple_{im}) \\ &= \mathbf{1}(\ln L_{im} - \ln(P + \delta Apple_{im}) + \beta' X_{im} + U_{im} > 0), \end{aligned}$$

where δ *captures the effect of Apple Music on consumer* i*'s valuation of the outside option. It is worth noting that in this example it is reasonable to claim that* $\mathrm{Cov}(Apple_{im}, U_{im}) = 0$ *because the launch date of Apple Music is unlikely to be correlated with individual heterogeneity.*

Case 2 (Switching Cost)**.** *The second example addresses the switching cost. Consider*

$$\begin{aligned} S_{im} &= \mathbf{1}[W_{im} > P - \delta \ln(1 + Tenure_{im})] \\ &= \mathbf{1}(\ln L_{im} - \ln(P - \delta \ln(1 + Tenure_{im})) + \beta' X_{im} + U_{im} > 0), \end{aligned}$$

where $Tenure_{im}$ *is consumer tenure up to the beginning of month* m*, and* $\delta > 0$*. For a new customer* i*, whose* $Tenure_{im} = 0$ *hence* $\ln(1+Tenure_{im}) = 0$*, the monetary cost of subscription is just the listed price* P*. For a current customer* i*, whose* $Tenure_{im} > 0$*, there is switching cost* $\delta \ln(1 + Tenure_{im})$ *involved in turning off the service. We use log transformation, so that the switching cost is concave in tenure. Note that in this example, it would be unreasonable to assume that* $\mathrm{Cov}(\ln(1 + Tenure_{im}), U_{im}) = 0$.

It is tempting to conclude that by using our previous results and the variation of X_{2im} (to identify δ in $P_{im} + \delta' X_{2im}$), we *can* identify parameters in both examples without price variation. Though this conjecture is correct under certain conditions (which could be strong in certain applications), it is incorrect in general. In general, we have Theorem 3, and the conclusion depends on whether or not X_{2im} and U_{im} are correlated. Because this extended model involves two new variables, P_{im} and X_{2im}, we need to rephrase the exogenous variation assumption and the normal distribution assumption.

Assumption 2′ (Exogenous Variation of Leisure and Price)**.** *Assume that* $(\mathbf{Z}_{im}, P_{im}) \perp\!\!\!\perp U_{im} \mid (X_{im}, \mu_i)$.

When there is no price variation $P_{im} = P$, the above assumption is the same as $\mathbf{Z}_{im} \perp\!\!\!\perp U_{im} \mid (X_{im}, \mu_i)$ in Assumption 2. It is worthwhile to understand how exogenous price variation provides additional information compared to the case with only usage variation. Note that we only seek to point out the additional information provided the exogenous price variation in addition to the usage variation. Our approach does not correct for the issues that arise due to endogeneity of prices. The latter has been extensively studied in the literature.[12]

Assumption 3′ (Normal Distribution—Extension)**.** *(1) For each month m, assume that*

$$U_{im} \mid (X_{im}, \mu_i) \sim \mathcal{N}(\sigma'_{u,x_2} X^*_{2im} + \sigma_{u,\mu}\mu^*_{im}, \sigma^2_{u2}),$$

where X^*_{2im} and μ^*_{im} are the residuals after applying the linear projection of X_{2im} and μ_i onto X_{1im}, respectively.

(2) Let $R_{im} \equiv (X'_{im}, \ln L_{im}, \mu_i)'$. Assume that $\mathrm{E}(R_{im}R'_{im})$ is of full rank.

Note that the rank condition in part (2) implies that X_{2im} cannot be a constant (recall that X_{im} already includes unit one), otherwise it can be shown that δ will not be identified.

Theorem 3 (Parametric Identification of WTP—Extension)**.** *Suppose Assumption 1, 2′, and 3′ hold. We have*

1. *(Case 1: X_{2im} and U_{im} are uncorrelated, i.e. $\sigma_{u,x_2} = 0$). All parameters β, δ, $\sigma_{u,\mu}$, and σ_{u2} are identified with or without price variation.*

[12] For example, we can use the control function approach to address endogeneity of price by letting X_{im} include the control variables for price (Petrin and Train, 2010).

2. *(Case 2: X_{2im} and U_{im} are correlated, i.e. $\sigma_{u,x_2} \neq 0$). All parameters β, δ, $\sigma_{u,\mu}$, σ_{u,x_2} and σ_{u2} are identified as long as we have at least* ***two*** *distinct prices. Without price variation, these parameters are poorly identified (see more discussion below).*

Following this theorem, we know that the model of Case 1, in which $X_{2im} = Apple_{im}$, is identified without price variation because $\text{Cov}(Apple_{im}, U_{im}) = 0$. In the second example, in which $X_{2im} = \ln(1 + Tenure_{im})$, it is unreasonable to claim $\text{Cov}(\ln(1 + Tenure_{im}), U_{im}) = 0$. Theorem 3 claims that this model will be poorly identified without price variation. It is shown in the proof of the above theorem that when there is no price variation, the identification depends on whether or not we can identify the parameters in the following nonlinear least square (NLS) regression:

$$Y_{im} = \ln(P + \delta' X_{2im}) - \psi_1 - \psi_2' X_{2im},$$

where Y_{im} is some *known* "dependent variable" defined in the proof. Note that the identification is possible only because $\ln(\cdot)$ is a nonlinear function. This kind of purely parametric identification can lead to poor estimation in practice because the log function is quite close to linear locally.[13] This raises serious concern about the collinearity between $\ln(1 + \delta' X_{2im}/P)$ and X_{2im}. This issue of poor identification is similar to the Heckman's two-step method for the sample selection model, in which the identification is possible only because the inverse Mills ratio is nonlinear (though it is close to linear). Having exogenous price variation resolves this difficulty (similar to the case in which Heckman's two-step method requires excluded variables that only affect selection but not the outcome). *Even with only two distinct prices, the theorem shows that we can identify the model.* Once the identification is clear, we estimate the model by the maximum likelihood estimator. Our simulation studies in the Online Appendix show that our estimator works well even with only two prices, and additional price variation (three distinct prices) does not bring noticeable efficiency gain.

[13] For example, note that $\ln(1 + c) \approx c$ when c is small. Define $\tilde{P} = P + \delta' \,\text{E}(X_{2im})$. We can write $\ln(P + \delta' X_{2im}) = \ln\left(1 + \frac{\delta'(X_{2im} - \text{E}(X_{2im}))}{P + \delta' \,\text{E}(X_{2im})}\right) + \ln(\tilde{P})$. Using $\ln(1 + c) \approx c$, we can see that $\ln(P + \delta' X_{2im})$ is also close to linear in $\delta'(X_{2im} - \text{E}(X_{2im}))$.

6 Empirical Application: Music Streaming Service

We focus on the market of online music streaming service in Southeast Asia during the period January 2016–December 2016. We represent the price in scaled $ term for exposition and to avoid attribution to the firm that provided the data. The usage (time of listening to music via this service) data are not scaled.

We examine an empirical setting in which we study the subscription decision of a customer. We use our method to obtain the estimates of the price elasticities of different segment of consumers (see results in Table 3), the revenue maximizing prices for each segment (in Table 3), and the distribution of the WTP for the monthly streaming service (Figure 2).

6.1 Data

The data were provided by a music streaming service company targeting the Southeast Asian market. Its service had 80% market share during the sample year. We will focus on the subscription choice of the monthly plan, and the price was always $149 for all consumers in our sample.

We observe the daily usage (the number of seconds each user listened to music with the service) of subscribers from January 1, 2016 to December 31, 2016. We also observe each user's payment transaction history during that period, so we observe consumer monthly subscription choices. In terms of demographics, we only observe age and gender. We sampled 300 users from one city, and found the daily weather information (precipitation and relative humidity) of that city during the sample period. These weather variables will be used as the exogenous variables that shift the daily leisure budget. All sampled users were subscribed to the streaming service in the first sample month (January 2016). At the end of our sample (December, 2016), 90% of the users were still subscribing to the service.

We have a few observations from the summary statistics detailed in Table 1. First, it is evident that the users who had cancelled their subscriptions at some point of time in our data used significantly less (less than one-half) than those who never cancelled the service. Second, younger and male users seem to be more likely to cancel their subscriptions. Third, consumers use the streaming music service less during weekends. This might be because weekends might involve other leisure activities, especially social

Table 1: Means of Key Variables in the Streaming Music Data (Jan 1, 2016–Dec 31, 2016)

	All Users	Never Cancelled	Ever Cancelled
Daily Usage (Hours)	1.37	1.45	0.61
	(2.26)	(2.33)	(1.33)
Daily Usage (Hours): Weekend	1.31	1.39	0.57
	(2.21)	(2.27)	(1.41)
Daily Usage (Hours): Weekdays	1.39	1.47	0.62
	(2.28)	(2.35)	(1.30)
Age	30.91	31.12	29.69
	(9.09)	(9.32)	(7.56)
Female (%)	42.00	42.35	40.00
Number of Users	300	255	45

Note: There is a single price ($149) for all consumers in the sample. The data are panel data at the daily frequency. The standard deviation is in the parenthesis.

activities.

6.2 Model

We need to specify the leisure equation, eq. (4), and the heterogeneous preference equation, eq. (6), for this particular application. First, let the daily leisure ℓ_{it} be

$$\ell_{it} = \begin{cases} \ell_{it}^* & \text{if } \ell_{it}^* > 0 \\ 0 & \text{if } \ell_{it}^* \leq 0, \end{cases}$$

$$\ell_{it}^* = \mu_i + \gamma_{i,Holiday} Holiday_t + \gamma_{i,Weekend} Weekend_t \\ + \gamma_{Precipitation} Precipitation_t + \gamma_{Humidity} Humidity_t + \varepsilon_{it}.$$

The exogenous variables Z_{it} in this application are $Precipitation_t$ and $Humidity_t$. $Holiday_t$ and $Weekend_t$ are dummy variables for holidays and weekends. Note that we also allow for heterogeneous effect of holidays and weekends. Second, we consider

Table 2: WTP for Music Streaming Service: Estimation Results

	Parameters	Estimates	Std Err
Usage eq.	$\mu_{Type\,1}$	-0.504	(0.049)
	$\mu_{Type\,2}$	0.666	(0.049)
	$\gamma_{Holiday,Type\,1}$	-0.005	(0.020)
	$\gamma_{Holiday,Type\,2}$	-0.113	(0.015)
	$\gamma_{Weekend,Type\,1}$	-0.008	(0.013)
	$\gamma_{Weekend,Type\,2}$	-0.055	(0.009)
	$\gamma_{Precipitation}$	-0.002	(0.000)
	$\gamma_{Humidity}$	0.003	(0.001)
Subscription eq.	β_0/σ_u	6.265	(1.970)
	$1/\sigma_u$	2.353	(0.792)
	β_{Age}/σ_u	0.065	(0.027)
	$\beta_{Age\,Sq}/\sigma_u$	-0.001	(0.000)
	β_{Female}/σ_u	0.189	(0.075)
	$\sigma_{u,\mu}/\sigma_u$	-2.487	(1.020)

Note: Two types of μ_i and the coefficients of holiday and weekend dummies were selected according to BIC. Seven types of r_{im} were selected according to BIC.

the linear projection of $\ln \alpha_{im}$ onto age, age squared, and the female gender indicator variable,

$$\ln \alpha_{im} = \beta_0 + \beta_{Age} Age_i + \beta_{Age\,Sq} Age_i^2 + \beta_{Female} Female_i + U_{im}.$$

6.3 Empirical Results

Table 2 presents the estimates of the main parameters of our model. From the estimates, we can see that the effect of weather on usage is at least statistically significant. Age has positive partial effect on WTP, and such an effect is declining as age grows. Women are willing to pay 18.9% more than men for this music streaming service.

Figure 2 plots the (unconditional) distribution function of the WTP for the subscription among all subscribers at the beginning of our sample period. The estimated

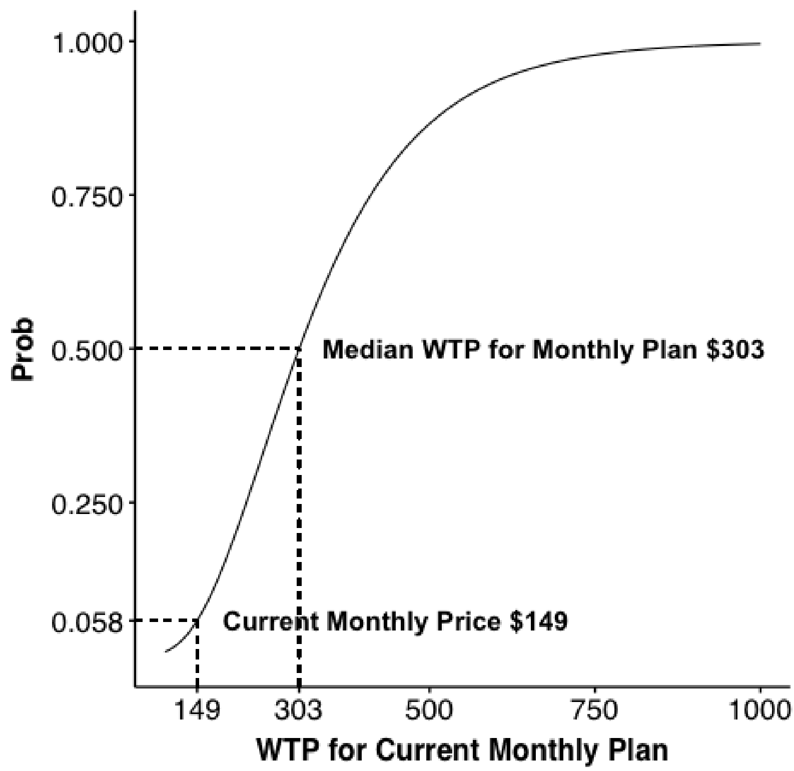

Figure 2: Estimates of the Distribution of WTP for the Monthly Plan

median WTP is \$303. According to the estimated distribution, only about 6% of current subscribers are willing to pay less than the listed price of \$149. This might explain the high market share and retention rate of this streaming service.

The model estimates can be connected to economically meaningful measures including price elasticities of different consumer segments and by computing the revenue maximizing prices. The price elasticity is defined as

$$e_{price} \equiv \frac{\partial \Pr(S_{im} = 1)}{\partial P} \frac{P}{\Pr(S_{im} = 1)},$$

and it is calculated using eq. (10), and its standard error was calculated using the delta method. For a monthly plan, consumers can always turn on and off the subscription—we do not consider the switching cost here because we have only one price. If the company wants to maximize the annual revenue (12 months), the revenue maximization

Table 3: Estimates of Price Elasticities, Median WTP and Revenue Maximizing Prices

Segment	Price Elasticity		Revenue Max Price	Mean Usage	Median WTP ($)
All Users	-0.27	(0.09)	226	1.37	303
Male	-0.30	(0.11)	220	1.43	296
Female	-0.24	(0.08)	234	1.29	313
Age ≤ 22	-0.33	(0.13)	213	1.45	287
Age 23–30	-0.30	(0.11)	220	1.55	295
Age > 30	-0.23	(0.07)	236	1.22	316

Note: "All Users" refer to the all sampled subscribers in Jan, 2016. The standard error of price elasticities estimates is in the parenthesis.

problem is the following,

$$\max_P \sum_{m=1}^{12} P(1 - F_{W,m}(P)).$$

Here $F_{W,m}(\cdot)$ is the distribution function of the WTP in month m, and $1 - F_{W,m}(P) = \Pr(W_{im} > P)$ is the percentage of consumers who will subscribe in month m. The distribution $F_{W,m}(\cdot)$ can vary month to month because the monthly leisure could change. The revenue maximizing monthly price satisfies

$$1 = \frac{P}{\sum_{m=1}^{12}(1 - F_{W,m}(P))} \sum_{m=1}^{12} \frac{\partial F_{W,m}(P)}{\partial P},$$

from which we can calculate the revenue maximizing price. Similarly, using the conditional distribution of the WTP given consumer demographics (age and gender), we can calculate the revenue maximizing monthly price if the company chooses to target specific consumer groups like student accounts in Spotify.

Table 3 reports the elasticities and revenue maximizing monthly prices. The estimates of price elasticities rephrase our earlier conclusion about WTP: younger people and men have higher price elasticities for this product. Overall, the subscribers are relatively inelastic suggesting that increasing price might be reasonable if the objective is to maximize the current revenue. According to our calculation, the revenue maximizing price will be $226 which is about 50 percent higher than the current price of $149. We also calculated the prices for other consumer segments. For example, the

Table 4: Estimates of Price Elasticities by Excluding One Weather Variable

User Groups	Humidity Only	Precipitation Only	Both
All Users	−0.235	−0.208	−0.274
	(0.090)	(0.089)	(0.094)
Male	−0.255	−0.226	−0.297
	(0.103)	(0.100)	(0.108)
Female	−0.207	−0.183	−0.241
	(0.076)	(0.075)	(0.077)
Age $\leq$ 22	−0.284	−0.252	−0.331
	(0.120)	(0.116)	(0.128)
Age 23–30	−0.259	−0.229	−0.302
	(0.105)	(0.102)	(0.111)
Age > 30	−0.199	−0.176	−0.231
	(0.071)	(0.070)	(0.071)

Note: The standard error is in the parenthesis. “All Users” refer to the all sampled subscribers in Jan, 2016 (the first month of our data).

revenue maximizing price for younger customers (age $\leq$ 22) who are usually students is \$213 which is 5% cheaper than our proposed regular price \$226.

When we compare usage and WTP across groups (the last two columns of Table 3), we have an interesting observation. Reading the column “Mean Usage”, we can see that women use less than men, and older consumers use less than younger users. Based on the usage pattern, one might think men and youths are willing to pay more for the subscription. Our estimates (the column “Median WTP (\$)”) show the opposite. This is because in our model, the WTP depends on both usage and the valuation of the leisure with the subscription. Even though women and older customers use less, they have higher valuation of the leisure as revealed by their higher subscription rate.

Lastly, one essential assumption is that the two weather variables create exogenous variation of usage/leisure, i.e. $\mathbf{Z}_{im} \perp\!\!\!\perp U_{im} \,|\, (X_{im}, \mu_i)$. Here $\mathbf{Z}_{im}$ consists of precipitation and humidity. If any of $\mathbf{Z}_{im}$ is correlated with U_{im}, the resulted estimates of WTP distribution (and other parameters, like price elasticities) will be biased. With two weather variables, we indeed over-identify our model—this belongs to the general over-identification issue of the generalized method of moments (GMM) model. So one way to check the exogenous assumption is to estimate the model using only one weather variable and to compare the estimates with the one using both weather variables. This practice has been used in Altonji, Elder and Taber (2005). In Table 4, we report the estimates of price elasticities using humidity or precipitation alone as the exogenous variation, and compare the estimates with the one using both weather variables. We do not observe substantial variation of the estimates suggesting that weather is a potentially exogenous factor.

7 Conclusion

Many subscription commerce markets charge the same price to every consumer and over time. Thus, price variation is very limited, and often non-existent. 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 high-frequency usage tracking data and observed subscription choices can identify the price elasticities and the distribution 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. with two price levels), can help identify more sophisticated models of WTP, including incorporating switching costs.

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 advance of technology like 5G telecommunications and the Internet of Things, the high-frequency 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.[14]

A Subscription Plan Examples

Table A1 details some common subscription services in the US. Some of the services are all inclusive with unlimited usage (e.g. Dropbox Premium) whereas others charge a marginal price for usage, or only include pre-specified quantities.

Table A1: 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.

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

B Nonparametric Identification and Estimation

In the basic model, we have the following subscription rule,

$$S_{im} = \mathbf{1}(\ln L_{im} - \ln P + \beta' X_{im} + U_{im} > 0),$$

where the expected monthly leisure L_{im} has been identified using the daily usage data. In this section, we will show that the exogenous usage variation ($\mathbf{Z}_{im}$) can identify the distribution of WTP without price variation even when we do not impose parametric assumptions about the joint distribution of (X_{im}, μ_i, U_{im}).

To state our result (the proof is in the Online Appendix), define the conditional choice probability (CCP) function,

$$\pi(x, \mu, l) \equiv \mathrm{E}(S_{im} \mid X_{im} = x, \mu_i = \mu, L_{im} = l).$$

Note that (a) $\pi(x, \mu, l)$ is nonparametrically estimable; (b) $\pi(x, \mu, l) = \Pr(S_{im} = 1 \mid X_{im} = x, \mu_i = \mu, L_{im} = l)$ by the binary nature of S_{im}.

Theorem B.1 (Nonparametric Identification and Estimation of WTP)**.** *Suppose Assumption 1 to 2 hold. We have that*

$$F_W(w \mid X_{im} = x, \mu_i = \mu, L_{im} = l) = 1 - \pi\left(x, \mu, \frac{P \times l}{w}\right),$$

provided that Pl/w is in the support of L_{im} conditional on (X_{im}, μ_i). In addition, if

(1) L_{im} is continuous,

(2) the support of P/L_{im} covers the support of α_{im} given X_{im} and μ_i,

(3) $\mathrm{E}(X_{im}X_{im}')$ is of full rank,

we have that

(1) the entire distribution $F_W(w \mid X_{im}, \mu_i, L_{im})$ is nonparametrically identified;

(2) the conditional mean of WTP equals,

$$\mathrm{E}(W_{im} \mid X_{im}, \mu_i) = \mathrm{E}(L_{im} \mid X_{im}, \mu_i)\, \mathrm{E}(Y_{1,im} \mid X_{im}, \mu_i),$$

where

$$Y_{1,im} = \frac{S_{im} - \mathbf{1}(L_{im} \geq \mathrm{E}(L_{im}))}{L_{im} f_L(L_{im} \mid X_{im}, \mu_i)} \frac{P}{L_{im}} - \frac{P}{\mathrm{E}(L_{im})},$$

and $f_L(L_{im} \mid X_{im}, \mu_i)$ is the conditional PDF of L_{im} given (X_{im}, μ_i);

(3) β can be consistently estimated by the OLS estimator

$$\hat{\beta} \equiv \left(\sum_{i=1}^{n} \sum_{m=1}^{M} X_{im} X'_{im} \right)^{-1} \left(\sum_{i=1}^{n} \sum_{m=1}^{M} X_{im} Y_{2,im} \right),$$

where

$$Y_{2,im} \equiv \frac{S_{im} - \mathbf{1}(\ln L_{im} \geq \mathrm{E}(\ln L_{im}))}{f_{\ln L}(\ln L_{im} \mid X_{im}, \mu_i)} + \mathrm{E}(\ln L_{im}) - \ln P.$$

where $f_{\ln L}(\cdot \mid X_{im}, \mu_i)$ is the conditional PDF of $\ln L_{im}$ given (X_{im}, μ_i).

The above theorem not only shows the identification of the WTP distribution, but also gives estimable formulas of the conditional distribution of W_{im}, and the conditional mean of WTP. The conditional mean can all be estimated by nonparametric regression easily. The support condition (the support of P/L_{im} covers the support of α_{im} given X_{im} and μ_i) can be restrictive when $\mathbf{Z}_{im}$ is discrete. If the support condition does not hold, we can use Theorem 2 that relies on the normal distribution assumption.

One way to check whether or not the support condition is going to hold is to leverage the parametric identification result. We want the support of P/L_{im} covers the support of α_{im} given X_{im} and μ_i. From data, we observe the range of P/L_{im} given (X_{im}, μ_i) because L_{im} has been estimated. Given the normal distribution assumption, we have

$$U_{im} \mid (X_{im}, \mu_i) \sim \mathcal{N}(\sigma_{u,\mu} \mu^*_{im}, \sigma_u^2).$$

Hence, $\alpha_{im} = \exp(X'_{im}\beta + U_{im})$ follows a log normal distribution given (X_{im}, μ_i). We then can compare the 95% confidence interval of α_{im} given (X_{im}, μ_i) with the observed range of P/L_{im} given (X_{im}, μ_i),

References

Altonji, Joseph G, and Richard K Mansfield. 2018. "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects." *American Economic Review*, 108(10): 2902–46.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources*, 40(4): 791–821.

Ariely, Dan. 2010. "Why Businesses Don't Experiment." *Harvard Business Review*, 88(4).

Bajari, Patrick, and C Lanier Benkard. 2005. "Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *Journal of Political Economy*, 113(6): 1239–1276.

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins. 2016. "A Dynamic Model of Demand for Houses and Neighborhoods." *Econometrica*, 84(3): 893–942.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics*, 25(2): 242–262.

Chamberlain, Gary. 1980. "Analysis of Covariance With Qualitative Data." *The Review of Economic Studies*, 47(1): 225–238.

Chen, Tony, Ken Fenyo, Sylvia Yang, and Jessica Zhang. 2018. "Thinking Inside the Subscription Box: New Research on E-Commerce Consumers." *McKinsey Quarterly*. Accessed: 2019-06-20.

Columbus, Louis. 2018. "The State of the Subscription Economy, 2018." *Forbes*. Accessed: 2019-06-20.

Danthurebandara, Vishva M, Jie Yu, and Martina Vandebroek. 2011. "Sequential Choice Designs to Estimate the Heterogeneity Distribution of Willingness-to-Pay." *Quantitative Marketing and Economics*, 9(4): 429–448.

Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior.* Cambridge University Press.

Diamond, Peter A, and Jerry A Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives*, 8(4): 45–64.

Ding, Min. 2007. "An Incentive-Aligned Mechanism for Conjoint Analysis." *Journal of Marketing Research*, 44(2): 214–223.

Gentzkow, Matthew. 2007. "Valuing New Goods in a Model With Complementarity: Online Newspapers." *American Economic Review*, 97(3): 713–744.

Green, Paul E, and Venkatachary Srinivasan. 1978. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research*, 5(2): 103–123.

Green, Paul E, and Vithala R Rao. 1971. "Conjoint Measurement for Quantifying Judgmental Data." *Journal of Marketing Research*, 8(3): 355–363.

Guadagni, Peter M, and John DC Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science*, 2(3): 203–238.

Hanemann, W Michael. 1994. "Valuing the Environment Through Contingent Valuation." *Journal of Economic Perspectives*, 8(4): 19–43.

Hausman, Jerry. 2012. "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives*, 26(4): 43–56.

Kalish, Shlomo, and Paul Nelson. 1991. "A Comparison of Ranking, Rating and Reservation Price Measurement in Conjoint Analysis." *Marketing Letters*, 2(4): 327–335.

Kamakura, Wagner A, and Gary J Russell. 1989. "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research*, 26(4): 379–390.

Lewbel, Arthur, Daniel McFadden, and Oliver Linton. 2011. "Estimating Features of a Distribution From Binomial Data." *Journal of Econometrics*, 162(2): 170–188.

Mitchell, Robert Cameron, and Richard 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, Aviv, John L Turner, and Jonathan W Williams. 2016. "Usage-Based Pricing and Demand for Residential Broadband." *Econometrica*, 84(2): 411–443.

Petrin, Amil, and Kenneth Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, 47(1): 3–13.

Rao, Vithala R. 2014. *Applied Conjoint Analysis.* Springer.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1): 34–55.

Stamatopoulos, Ioannis, Achal Bassamboo, and Antonio Moreno. 2021. "The Effects of Menu Costs on Retail Performance: Evidence From Adoption of the Electronic Shelf Label Technology." *Management Science*, 67(1): 242–256.

Train, Kenneth, and Melvyn Weeks. 2005. "Discrete Choice Models in Preference Space and Willingness-to-Pay Space." In *Applications of Simulation Methods in Environmental and Resource Economics*. 1–16. Springer.

Varian, Hal R. 1992. *Microeconomic Analysis.* . 3rd ed., New York:W. W. Norton & Company, Inc.

Voelckner, Franziska. 2006. "An Empirical Comparison of Methods for Measuring Consumers' Willingness to Pay." *Marketing Letters*, 17(2): 137–149.

Wertenbroch, Klaus, and Bernd Skiera. 2002. "Measuring Consumers' Willingness to Pay at the Point of Purchase." *Journal of Marketing Research*, 39(2): 228–241.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* . 2nd ed., Cambridge, Massachusetts:The MIT Press.

Xie, Jinhong, and Steven M Shugan. 2001. "Electronic Tickets, Smart Cards, and Online Prepayments: When and How to Advance Sell." *Marketing Science*, 20(3): 219–243.

Supplement for

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

Cheng Chou [15] Vineet Kumar [16]

November 2021

A Proofs of the Theorems in the Paper

Proof of Theorem 1. When the utility function $u_{im(t)}(q_{it}, q_{0,it})$ is homogeneous of degree 1, it is a homothetic utility. Let Q_{it} and $Q_{0,it}$ denote the optimal solutions of q_{it} and $q_{0,it}$ that maximize $u_{im(t)}(q_{it}, q_{0,it})$ subject to $q_{it} + q_{0,it} = \ell_{it}$. These optimal solutions are also the observed usage by our assumption. The demand functions satisfy (e.g. page 147 of Varian, 1992)

$$Q_{it} = r_{im(t)}\ell_{it}, \qquad \text{and} \qquad Q_{0it} = r_{0,im(t)}\ell_{it},$$

for some parameter $r_{im(t)}$, and $r_{0,im(t)} = 1 - r_{im(t)}$ by the budget constraint. This is part (2) of Theorem 1.

In order to show part (1) of this theorem. We first derive the the indirect utility on day t, denoted by V_{it}, in month m. When the monthly utility is additively separable in daily utilities, the monthly utility for month m equals $\sum_{t:m(t)=m} V_{it}$. The daily indirect utility V_{it} equals the following,

$$\begin{aligned}
V_{it} &\equiv u_{im(t)}(Q_{it}, Q_{0,it}) \\
&= Q_{it} u_{im(t)}(1, Q_{0,it}/Q_{it}) && \text{as } u_{im(t)} \text{ is homogeneous of degree 1} \\
&= Q_{it} u_{im(t)}(1, r_{0,im(t)}/r_{im(t)}) && \text{by } Q_{0,it}/Q_{it} = r_{0,im(t)}/r_{im(t)} \\
&= \ell_{it} \underbrace{r_{im(t)} u_{im(t)}(1, r_{0,im(t)}/r_{im(t)})}_{\equiv \alpha_{im(t)}} && \text{by } Q_{it} = r_{im(t)}\ell_{it}.
\end{aligned}$$

[15] Cheng Chou: School of Business, University of Leicester, UK. Email: cchou@le.ac.uk.
[16] Vineet Kumar: School of Management, Yale University `email:`derdenge@andrew.cmu.edu.

We conclude that the daily indirect utility is $V_{it} = \alpha_{im(t)}\ell_{it}$, hence the monthly indirect utility for month m equals

$$\sum_{t:m(t)=m} V_{it} = \alpha_{im} \sum_{t:m(t)=m} \ell_{it}.$$

Next, we need to compute the expected monthly indirect utility. Let I_{im} denote the set of information available at the beginning of month m according to Assumption 1. The expected indirect utility of month m is

$$W_{im} = \alpha_{im} \underbrace{\sum_{t:m(t)=m} \mathrm{E}(\ell_{it} \mid I_{im})}_{\equiv L_{im}}.$$

We now derive the formula of the expected monthly leisure L_{im} using the properties of a Tobit model. Because ℓ_{it} defined in eq. (4) follows a Tobit model, we have

$$\begin{aligned}\mathrm{E}(\ell_{it} \mid I_{im}) &= \mathrm{E}(1(\ell^*_{it} > 0)\ell^*_{it} \mid I_{im}) \\ &= \Pr(\ell^*_{it} > 0 \mid I_{im})\, \mathrm{E}(\ell^*_{it} \mid I_{im}, \ell^*_{it} > 0).\end{aligned}$$

We first have

$$\Pr(\ell^*_{it} > 0 \mid I_{im}) = \Pr(\varepsilon_{it} > -(\mu_i + \gamma' Z_{it})) = \Phi(\mu_i + \gamma' Z_{it}).$$

Secondly, we have

$$\begin{aligned}\mathrm{E}(\ell^*_{it} \mid I_{im}, \ell^*_{it} > 0) &= \mathrm{E}(\ell^*_{it} \mid I_{im}, \ell^*_{it} > 0) \\ &= \mu_i + \gamma' Z_{it} + \frac{\phi(\mu_i + \gamma' Z_{it})}{\Phi(\mu_i + \gamma' Z_{it})} \quad \text{by the property of Tobit models.}\end{aligned}$$

Here we used the property of a Tobit model (see eq. (17.10) of Wooldridge (2010)). Hence we have that

$$\mathrm{E}(\ell_{it} \mid I_{im}) = \left[\left(\mu_i + \gamma' Z_{it} + \frac{\phi(\mu_i + \gamma' Z_{it})}{\Phi(\mu_i + \gamma' Z_{it})}\right) \Phi(\mu_i + \gamma' Z_{it})\right].$$

Then L_{im} takes the form of eq. (5), and part (1) of this theorem has been proved. □

For simplicity, in the rest of the proofs, we omit the subscript "$im(t)$", "im", and "it" whenever there is no confusion.

Proof of Theorem 2. We first show that the vector of unknown parameters $(\beta', \sigma_u, \sigma_{u,\mu})'$ is identified. Then we show that the identification of $(\beta', \sigma_u, \sigma_{u,\mu})'$ implies the identification of the distribution of WTP.

We have that

$$\begin{aligned}\Pr(S = 1 \mid X, \mu, L) &= \Pr(U > \ln P - \beta' X - \ln L \mid X, \mu, L) \\ &= \Pr(U > \ln P - \beta' X - \ln L \mid X, \mu),\end{aligned}$$

where the second line follows from $\mathbf{Z} \perp\!\!\!\perp U \mid (X, \mu)$ hence $L \perp\!\!\!\perp U \mid (X, \mu)$. Now by assumption 3, we have

$$U \mid (X, \mu) \sim \mathcal{N}(\sigma_{u,\mu}\mu^*,\ \sigma_u^2). \tag{A.1}$$

It then follows from eq. (A.1) that

$$\begin{aligned}\Pr(S = 1 \mid X, \mu, L) &= 1 - \Phi\left(\sigma_u^{-1}\Big[\ln P - \beta' X - \ln L - \sigma_{u,\mu}\mu^*\Big]\right) \\ &= \Phi\left(\sigma_u^{-1}\Big[\ln L - \ln P + \beta' X + \sigma_{u,\mu}\mu^*\Big]\right).\end{aligned} \tag{A.2}$$

To see the identification, note two things. First, $\Pr(S = 1 \mid X, \mu, L)$ is observable from data—μ and L have been identified before using usage data. Second, the CDF Φ is strictly increasing. We then have that

$$\begin{aligned}\Phi^{-1}(\Pr(S = 1 \mid X, \mu, L)) &= \sigma_u^{-1}\Big[\ln L - \ln P + \beta' X + \sigma_{u,\mu}\mu^*\Big] \\ &= \sigma_u^{-1}(\beta_0 - \ln P) + (\sigma_u^{-1}\beta_1)' X_1 + \sigma_u^{-1}\ln L + (\sigma_u^{-1}\sigma_{u,\mu})\mu^*.\end{aligned}$$

The second line follows from $\beta' X = \beta_0 + \beta_1' X_1$. We want to rewrite the above equation as a "linear regression". Define the vector of "regressors" $\tilde{X} \equiv (1, X_1', \ln L, \mu^*)'$, and define the vector of "regression coefficients" $\tilde{\beta} \equiv (\sigma_u^{-1}(\beta_0 - \ln P), \sigma_u^{-1}\beta_1', \sigma_u^{-1}, \sigma_u^{-1}\sigma_{u,\mu})'$. Note that the identification of $\tilde{\beta}$ implies the identification of β, σ_u, and $\sigma_{u,\mu}$. Using the new notation, the above display reads

$$\Phi^{-1}(\Pr(S = 1 \mid X, \mu, L)) = \tilde{X}'\tilde{\beta}. \tag{A.3}$$

Equation (A.3) resembles a linear regression. Multiplying both sides of eq. (A.3) by the vector $\tilde{X}$ and taking expectation, we have

$$\mathrm{E}\Big[\tilde{X}\Phi^{-1}(\Pr(S = 1 \mid \mu_i, \mathbf{Z}, X))\Big] = \mathrm{E}(\tilde{X}\tilde{X}')\tilde{\beta},$$

So $\tilde{\beta}$ is identified, if $\mathrm{E}(\tilde{X}\tilde{X}')$ is invertible. Because μ^* is a linear combination of X_1 and μ, $\mathrm{E}(\tilde{X}\tilde{X}')$ is invertible if and only if $\mathrm{E}(RR')$ is of full rank, where $R \equiv (1, X_1', \ln L, \mu)'$, as assumed in Assumption 3.

Second, we derive the distribution of WTP. We have that

$$\begin{aligned} F_W(w \mid X, \mu, L) &= \Pr(\ln W \leq \ln w \mid X, \mu, L) \\ &= \Pr(\ln L + \beta' X + U \leq \ln w \mid X, \mu, L) \\ &= \Pr(U \leq \ln w - (\ln L + \beta' X) \mid X, \mu) && \text{by } U \perp\!\!\!\perp L \mid (X, \mu) \\ &= \Phi(\sigma_u^{-1}[\ln w - (\ln L + \beta' X) - \sigma_{u,\mu}\mu^*]) && \text{by eq. (A.1).} \end{aligned}$$

□

Proof of Theorem 3. In this proof, it is convenient to write $g(P, X_2; \delta) \equiv \ln(P + \delta' X_2)$, or simply g when there is no confusion. We have that

$$\begin{aligned} \Pr(S = 1 \mid \mu, L, X, P) &= \Pr(U > g - \beta_0 - \beta_1' X_1 - \ln L \mid \mu, L, X, P) \\ &= \Pr(U > g - \beta_0 - \beta_1' X_1 - \ln L \mid X, \mu), \end{aligned}$$

where the second line follows from $(\mathbf{Z}, P) \perp\!\!\!\perp U \mid (X, \mu)$, hence $(L, P) \perp\!\!\!\perp U \mid (X, \mu)$. By assumption $3'$, we have

$$U \mid (X, \mu) \sim \mathcal{N}\left(\sigma_{u,\mu}\mu^* + \sigma_{u,x_2}' X_2^*,\ \sigma_{u2}^2\right). \tag{A.4}$$

It follows from eq. (A.4) that

$$\Pr(S = 1 \mid \mu, L, X, P) = \Phi\left(\sigma_{u2}^{-1}\left[(\beta_0 - g) + \beta_1' X_1 + \ln L + \sigma_{u,\mu}\mu^* + \sigma_{u,x_2}' X_2^*\right]\right). \tag{A.5}$$

By the invertibility of the CDF Φ, we have

$$\Phi^{-1}(\Pr(S = 1 | \mu, L, X, P)) = \sigma_{u2}^{-1}(\beta_0 - g) + (\sigma_{u2}^{-1}\beta_1)' X_1 + \sigma_{u2}^{-1} \ln L + (\sigma_{u2}^{-1}\sigma_{u,\mu})\mu^* + (\sigma_{u2}^{-1}\sigma_{u,x_2})' X_2^*.$$

We know $\Phi^{-1}(\Pr(S = 1 \mid \mu, L, X, P))$ from data. This above display resembles a NLS regression model. The nonlinearity comes from g function. The condition whether or not σ_{u,x_2} equals zero matters for the rest of arguments. We first consider the case when $\sigma_{u,x_2} = 0$, which is simpler, then proceed to the case when $\sigma_{u,x_2} \neq 0$.

Case 1: $\sigma_{u,x_2} = 0$**.** If $\sigma_{u,x_2} = 0$, we have

$$\Phi^{-1}(\Pr(S = 1 \mid \mu, L, X, P)) = \sigma_{u2}^{-1} \underbrace{(\beta_0 - g)}_{\equiv H_1(X_2)} + (\sigma_{u2}^{-1}\beta_1)' X_1 + \sigma_{u2}^{-1} \ln L + (\sigma_{u2}^{-1}\sigma_{u,\mu})\mu^*.$$

For any fixed X_2, the above resembles a linear regression because $H_1(X_2)$ is a constant for a given X_2. When X_2 is fixed, $\sigma_{u2}^{-1} H_1(X_2)$ can be viewed as intercept term. For each value of X_2, define $\tilde{X} \equiv (1, X_1', \ln L, \mu^*)'$. If for any value of X_2, we have that $\mathrm{E}[\tilde{X}(\tilde{X})']$ has full rank, we can identify $\sigma_{u2}^{-1} H_1(X_2)$, σ_{u2}^{-1}, $\sigma_{u2}^{-1}\beta_1$, $\sigma_{u2}^{-1}\sigma_{u,\mu}$, hence $H_1(X_2)$. We then have identified β_1, $\sigma_{u\mu}$ and σ_{u2} by transformation. Note that μ^* is the residual of the linear projection of μ onto X_1, hence it is a linear combination of μ and X_1. So if $\mathrm{E}(RR')$ has full rank with $R = (1, \ln L, X_1', \mu)'$ (as assumed in Assumption 3′), we have $\mathrm{E}[\tilde{X}(\tilde{X})']$ has full rank.

We now knows $H_1(X_2)$ for any value of X_2, and

$$H_1(X_2) = \beta_0 - g.$$

This is just a NLS regression. Now recall $g(P_{im}, X_{2im}; \delta) = \ln(P_{im} + \delta' X_{2im})$, we have

$$H_1(X_{2im}) = \beta_0 - \ln(P_{im} + \delta' X_{2im}) \tag{A.6}$$

for any value of X_{2im}. The objective is to solve (β_0, δ) from eq. (A.6).

Here we only prove the simple case when X_{2im} is a scalar as in our two examples. The general case when X_{2im} can be shown very similarly. Denote X_{2a} and X_{2b} two distinct values that X_{2im} can take (e.g. $X_{2a} = 1$ and $X_{2b} = 0$ in Case 1),[17] and let $H_{1a} = H_1(X_{2a})$ and $H_{1b} = H_1(X_{2b})$. Subtracting eq. (A.6) when $X_{2im} = X_{2a}$ and X_{2b}, we have

$$H_{1b} - H_{1a} = \ln\left(\frac{P + \delta X_{2a}}{P + \delta X_{2b}}\right) = \ln\left[1 + \frac{\delta}{P + \delta X_{2b}}(X_{2a} - X_{2b})\right].$$

For a fixed X_{2b},[18] viewing

$$\tilde{\delta} \equiv \frac{\delta}{P + \delta X_{2b}}$$

as an unknown coefficient, we have a linear equation of $\tilde{\delta}$,

$$(\exp(H_{1b} - H_{1a}) - 1) = \tilde{\delta}(X_{2a} - X_{2b}).$$

[17] X_{2im} at least takes two values, otherwise $\mathrm{E}[(1, X_2')'(1, X_2')]$ does not have full rank as assumed in Assumption 3′.

[18] If 0 is in the support X_{2im}, letting $X_{2b} = 0$ gives rise to $\tilde{\delta} = \delta/P$.

This gives rise to a solution of $\tilde{\delta}$. Knowing $\tilde{\delta}$, δ can be solved if we know δX_{2b}. It is known by noting that

$$\tilde{\delta} X_{2b} = \frac{\delta X_{2b}}{P + \delta X_{2b}}.$$

We then have

$$\delta X_{2b} = \frac{\tilde{\delta} X_{2b} P}{1 - \tilde{\delta} X_{2b}} \qquad \text{and} \qquad \delta = \tilde{\delta}\left(P + \frac{\tilde{\delta} X_{2b} P}{1 - \tilde{\delta} X_{2b}}\right)$$

After obtaining δ, we can solve $\beta_0 = \ln(P + \delta' X_{2im}) + H_1(X_{2im})$. Note that the above arguments do not require price variation.

Case 2: $\sigma_{u,x_2} \neq 0$. When $\sigma_{u,x_2} \neq 0$, the above arguments do not proceed. Particularly, for a fixed X_2, X_2^* and μ^* are linear combination of X_1 and μ. Hence we have collinearity issue in the regression. To clarify the issue, it helps express μ^* and X_2^* explicitly:

$$\begin{aligned} \mu^* &= \mu - \mathrm{E}(\mu) + A_\mu(X_1 - \mathrm{E}(X_1)) \\ X_2^* &= X_2 - \mathrm{E}(X_2) + A_{x2}(X_1 - \mathrm{E}(X_1)), \end{aligned}$$

with

$$A_\mu = \mathrm{Cov}(\mu, X_1)[\mathrm{Var}(X_1)]^{-1}, \qquad \text{and} \qquad A_{x2} = \mathrm{Cov}(X_2, X_1)[\mathrm{Var}(X_1)]^{-1}.$$

Note that both A_μ and A_{x2} are identified from data. Standard but tedious calculation reveals that

$$\begin{aligned} \Phi^{-1}(\Pr(S = 1 \mid \mu, L, X, P)) &= \sigma_{u2}^{-1} \underbrace{(\beta_0 - g + \beta_1' \mathrm{E}(X_1) + \sigma_{u,x_2}'(X_2 + \mathrm{E}(X_2)))}_{\equiv H_2(P, X_2)} \\ &\quad + \sigma_{u2}^{-1} \ln L + \sigma_{u2}^{-1} \underbrace{(\beta_1' + \sigma_{u,\mu} A_\mu + \sigma_{u,x_2}' A_{x2})}_{\equiv \psi_1'}(X_1 - \mathrm{E}(X_1)) + \sigma_{u2}^{-1} \sigma_{u,\mu}(\mu - \mathrm{E}(\mu)). \end{aligned}$$

Now for a fixed X_2, the above resembles a linear regression. If $\mathrm{E}(RR')$ has full rank, we can identify σ_{u2}^{-1}, ψ_1, and $\sigma_{u,\mu}$. If we could further identify σ_{u,x_2}, we can obtain β_1 from ψ_1.

We proceed to leverage to the NLS regression:

$$H_2(P, X_2) = \tilde{\beta}_0 + \sigma_{u,x_2}' X_2 - g, \tag{A.7}$$

where

$$\tilde{\beta}_0 \equiv \beta_0 + \beta_1' \operatorname{E}(X_1) - \sigma_{u,x_2}' \operatorname{E}(X_2).$$

The essence of identification now resides at the identification of δ, β_0 and σ_{u,x_2} from eq. (A.7). Now letting $g(P_{im}, X_2; \delta) = \ln(P_{im} + \delta' X_2)$, we have

$$H_2(P, X_2) = \tilde{\beta}_0 + \sigma_{u,x_2}' X_2 - \ln(P_{im} + \delta' X_2). \tag{A.8}$$

If we observe at least two prices, P_a and P_b, we have

$$\begin{aligned} H_2(P_a, X_2) &= \tilde{\beta}_0 + \sigma_{u,x_2}' X_2 - \ln(P_a + \delta' X_2) \\ H_2(P_b, X_2) &= \tilde{\beta}_0 + \sigma_{u,x_2}' X_2 - \ln(P_b + \delta' X_2). \end{aligned}$$

Subtracting one from the other equation, we have

$$\exp(H_2(P_b, X_2) - H_2(P_a, X_2)) - 1 = \frac{(P_a - P_b) + \delta' X_2}{P_b + \delta' X_2},$$

which can be viewed as an equation of the unknown $\delta' X_2$. We can rewrite this equation as follows,

$$H_3(X_2) = \delta' X_2,$$

where

$$H_3(X_2) = \frac{[\exp(H_2(P_b, X_2) - H_2(P_a, X_2)) - 1]P_b + (P_b - P_a)}{1 - [\exp(H_2(P_b, X_2) - H_2(P_a, X_2)) - 1]}.$$

Hence if $\operatorname{E}(X_2 X_2')$ is of full rank, we can identify δ with only two prices. Once we know δ, we can rearrange eq. (A.8)

$$\ln(P_{im} + \delta' X_2) + H_2(P, X_2) = \tilde{\beta}_0 + \sigma_{u,x_2}' X_2.$$

The left-hand-side variable is now observable, hence $\tilde{\beta}_0$, and σ_{u,x_2} is identified when $\operatorname{E}[(1, X_2')'(1, X_2')]$ has full rank. Without price variation, the regression form eq. (A.8) suffers for serious collinearity because $\ln(P + \delta' X_2)$ and X_2 are highly collinear. $\square$

Proof of Theorem B.1. The starting point is the conclusion of Theorem 1:

$$W = \alpha L.$$

Using the subscription decision rule,

$$S = \mathbf{1}(\alpha L > P) = \mathbf{1}\left(-\frac{P}{L} + \alpha > 0\right)$$

It will be convenient to denote

$$V^* = -\frac{P}{L},$$

and write

$$S = \mathbf{1}(V^* + \alpha > 0).$$

Because L is a function of $\mathbf{Z}$ and μ, V^* is also a function of μ_i and $\mathbf{Z}$ for a fixed price P. Assumption 2 ($\mathbf{Z} \perp\!\!\!\perp U \mid (X, \mu)$) implies $\alpha \perp\!\!\!\perp V^* \mid (X, \mu)$ because α is a function of X and U.

First, we derive $F_W(w \mid X, \mu, L)$. We have that

$$\begin{aligned}
F_W(w \mid X, \mu, L) &= \Pr(W \leq w \mid X, \mu, L) \\
&= \Pr(\alpha L \leq w \mid X, \mu, L) \quad \text{by } L > 0 \\
&= \Pr\Big(\alpha \leq \frac{w}{L} \,\Big|\, X, \mu, L\Big) \\
&= \Pr\Big(\alpha \leq \frac{w}{L} \,\Big|\, X, \mu\Big) \qquad \text{by } U \perp\!\!\!\perp \mathbf{Z} \mid (X, \mu), \text{ hence } \alpha \perp\!\!\!\perp L \mid (X, \mu).
\end{aligned}$$

The objective now is to obtain the distribution of $F_\alpha(a \mid X, \mu)$. We have that

$$\begin{aligned}
\pi(x, \mu, l) &= \Pr(S = 1 \mid X = x, \mu_i = \mu, L = l) \\
&= \Pr(\alpha > -v^* \mid X = x, \mu_i = \mu, L = l) \qquad\qquad \text{where } v^* = -P/l \\
&= \Pr(\alpha > -v^* \mid X = x, \mu_i = \mu) \\
&= 1 - F_\alpha(-v^* \mid X = x, \mu_i = \mu).
\end{aligned}$$

Alternatively, we can write

$$1 - \pi\left(x, \mu, \frac{P}{a}\right) = F_\alpha(a \mid X = x, \mu_i = \mu),$$

provided that P/a is in the support of L. Now we return to the question: what is $\Pr(\alpha \leq w/L \mid X, \mu)$? We have that

$$\begin{aligned}
\Pr\Big(\alpha \leq \frac{w}{L} \,\Big|\, X, \mu\Big) &= F_\alpha\left(\frac{w}{L} \mid X, \mu\right) \\
&= 1 - \pi\left(X, \mu_i, \frac{PL}{w}\right).
\end{aligned}$$

We have the conclusion that the conditional CDF of WTP is

$$F_W(w \mid X = x, \mu_i = \mu, L = l) = 1 - \pi\left(x, \mu, \frac{Pl}{w}\right),$$

provided that $(Pl)/w$ is the in support of L conditional of (X, μ).

Second, we derive the formula of $\mathrm{E}(W \mid X, \mu)$. By the conditional independence Assumption 2, we have

$$\mathrm{E}(W \mid X, \mu) = \mathrm{E}(L \mid X, \mu)\,\mathrm{E}(\alpha \mid X, \mu).$$

Because we observe (L, X, μ), we only need to identify $\mathrm{E}(\alpha \mid X, \mu)$. Using the similar arguments in Lewbel (2014), it can be shown that

$$\mathrm{E}(\alpha \mid X, \mu) = \mathrm{E}(Y_1 \mid X, \mu),$$

where

$$Y_1 = \frac{S - \mathbf{1}(V^* \geq \mathrm{E}(V^*))}{f_{V^*}(V^* \mid X, \mu)} + \mathrm{E}(V^*),$$

if V^* is continuous and that the support of V^* covers the support of α given (X, μ). We can rewrite Y_1 equivalently as follows,

$$Y_1 = \frac{S - \mathbf{1}(L \geq \mathrm{E}(L))}{L f_L(L \mid X, \mu)}\frac{P}{L} - \frac{P}{\mathrm{E}(L)},$$

where f_L denotes the observable conditional distribution of L given (X, μ). To see this, first note

$$\mathbf{1}(V^* \geq \mathrm{E}(V^*)) = \mathbf{1}(L \geq \mathrm{E}(L)).$$

Next, viewing L as a transformation of $V^* = -P/L$, we have that

$$f_{V^*}(V^* \mid X, \mu) = f_L(L \mid X, \mu)\frac{L^2}{P}.$$

In summary, we have

$$\mathrm{E}(W \mid X, \mu) = \mathrm{E}(L \mid X, \mu)\,\mathrm{E}(Y_1 \mid X, \mu).$$

Third, we show the OLS estimator of β. We can write the subscription decision rule as follows,

$$S = \mathbf{1}\left(\ln(L/P) + \beta' X + U > 0\right).$$

Note that $\ln(L/P) \perp\!\!\!\perp (\beta' X + U) \mid (X, \mu)$ because L is a function of $(\mu, \mathbf{Z})$, and $\mathbf{Z} \perp\!\!\!\perp U \mid (X, \mu)$ as stated in Assumption 2. Using similar arguments in section 6 of Lewbel (2014), it can be shown that

$$\mathrm{E}(Y_2 \mid X, \mu) = \mathrm{E}(X'\beta + U \mid X, \mu).$$

Multiplying both sides of the above display by X, we have that

$$\begin{aligned}\mathrm{E}(XY \mid X, \mu) &= \mathrm{E}(XX'\beta + XU \mid X, \mu) \\ &= XX'\beta + \mathrm{E}(XU \mid X, \mu).\end{aligned}$$

Taking unconditional expectation of both sides of the above display, we have

$$\mathrm{E}(XY) = \mathrm{E}(XX')\beta + \mathrm{E}(\mathrm{E}(XU \mid X, \mu)) = \mathrm{E}(XX')\beta,$$

because

$$\mathrm{E}(\mathrm{E}(XU \mid X, \mu)) = \mathrm{E}(XU) = 0.$$

Thus, we can express $\beta = \mathrm{E}(XX')^{-1}\,\mathrm{E}(XY_2)$ when $\mathrm{E}(XX')$ is of full rank. □

B Simulation Studies Based on Real Data

B.1 Data Generation Process

We design a simulation study based on two individual level survey data (the Dutch Time Use Survey (DTUS) 2005, and the Living Costs and Food Survey (LCF) 2018–2019 in the UK) to make our data generation process closer to the reality. To generate data, we need to specify the leisure equation that will give us the usage, and the subscription choice equation. The context in this simulation is household subscription and usage (hours of watching) of TV streaming services like the Netflix. From the DTUS, we can observe how people spent time in many activities including watching TV along with demographic variables. The specification of our usage and leisure equation is based on the DTUS data. To specify the subscription choice equation, we explore the LCF data, which provides microdata about household expenditure on digital or online entertainment subscription(s) such as Netflix and household characteristics.

First, we specify the usage and leisure process based on the DTUS data:

$$Q_{it} = r_i \ell_{it}, \tag{B.1a}$$

$$\ell_{it} = \begin{cases} \ell^*_{it} & \text{if } \ell^*_{it} > 0 \\ 0 & \text{if } \ell^*_{it} \leq 0, \end{cases} \tag{B.1b}$$

$$\ell^*_{it} = \mu_i + \gamma_1 Z_{it} + \gamma_{Mon} Mon_t + \gamma_{Tue} Tue_t + \cdots + \gamma_{Sat} Sat_t + \varepsilon_{it}. \tag{B.1c}$$

Here $Mon_t, \dots, Sat_t$ are the day dummy variables controlling seasonality. We determine consumer fixed effect μ_i by his/her demographic information in the DTUS, and Z_{it} is a variable that affects the observed leisure other than μ_i.[19] Using the observed percentage of leisure time spent in watching TV in the DTUS, we obtain r_i for each individual i. By the K-means classification, we observe four types of (μ_i, r_i) in the sample, and they are more or less evenly distributed. We let $\varepsilon_{it} \sim \mathcal{N}(0, 1)$. Note that ℓ_{it} in the sample is measured in hours, so the unit standard deviation of ε_{it} is realistic. The true values of all parameters are in the tables of results that will be discussed later.

Second, we specify the subscription choice equation based on the LCF data. There are two versions below depending on whether or not there is price variation. The first one does not have price variation, and it is

$$\begin{aligned} S_{im} = \mathbf{1}(&\ln L_{im} + \beta_0 + \beta_{Age} Age_i + \beta_{Hsize} Hsize_i \\ &+ \beta_{MiddleIncome} MiddleIncome_i + \beta_{HighIncome} HighIncome_i - \ln P + U_{im}). \end{aligned} \tag{B.2}$$

$Hsize$ is the household size, and $MiddleIncome$ and $HighIncome$ are the dummy variables of income level. The expected monthly leisure L_{im} is computed from the leisure eq. (B.1). The error term $U_{im} \sim \mathcal{N}(0, 1)$, and it is correlated with μ_i. We let price $P = 28$, so that the resulted proportion of subscription in our simulation is about 30%, which is close to the actual proportion of subscription in the UK (24%). See the tables below for the values of other parameters.

In the second version of the subscription choice equation, we consider the case with price variation using the example of switching cost (Case 2)

$$\begin{aligned} S_{im} = \mathbf{1}(&\ln L_{im} + \beta_0 + \beta_{Age} Age_i + \beta_{Hsize} Hsize_i + \beta_{MiddleIncome} MiddleIncome_i \\ &+ \beta_{HighIncome} HighIncome_i - \ln[P + 28 \times \delta_{Tenure} \ln(1 + Tenure_{im})] + U_{im}). \end{aligned} \tag{B.3}$$

The tenure in our simulation is the number of months with active subscription. The initial tenure was randomly drawn from 0, 1, …, 3 with even probabilities. The tenure is correlated with the unobserved heterogeneity U_{im}.

[19] In the DTUS, we observe leisure, age, household size, income and the number of children. We first run a linear regression of $\ell_{it} = \mu_i^* + \gamma_{Mon} Mon_t + \gamma_{Tue} Tue_t + \cdots + \gamma_{Sat} Sat_t + \varepsilon_{it}$ with individual specific intercept μ_i^*. Then we regress the estimates of μ_i^* on age, household size, income and the number of children. Let μ_i be the fitted value of this regression, and let Z_i^* be the residuals. Finally, we generate $Z_{it} = Z_i^* + \omega_{it}$ with $\omega_{it} \sim \mathcal{N}(0, 1)$.

Table 2: Simulation without Price Variation

		Estimates: Single Price					
Parameters	Truth	$n = 1e3, M = 10$		$n = 2e3, M = 10$		$n = 2e3, M = 5$	
(I) Usage eq. (B.1)							
γ_1	1.000	1.004	(0.005)	1.002	(0.003)	1.002	(0.004)
γ_{Mon}	−4.960	−4.979	(0.025)	−4.973	(0.017)	−4.969	(0.022)
γ_{Tue}	−3.550	−3.562	(0.021)	−3.557	(0.014)	−3.553	(0.018)
γ_{Wed}	−2.720	−2.726	(0.020)	−2.723	(0.013)	−2.721	(0.017)
γ_{Thu}	−3.390	−3.402	(0.020)	−3.396	(0.014)	−3.395	(0.019)
γ_{Fri}	−1.610	−1.619	(0.018)	−1.618	(0.012)	−1.616	(0.016)
γ_{Sat}	2.350	2.355	(0.018)	2.349	(0.012)	2.350	(0.017)
$\mu_{Type\,1}$	8.000	8.044	(0.071)	8.024	(0.048)	8.016	(0.070)
$r_{Type\,1}$	0.100	0.100	(0.001)	0.100	(0.001)	0.100	(0.001)
$\mu_{Type\,2}$	10.000	10.012	(0.040)	10.010	(0.029)	10.007	(0.036)
$r_{Type\,2}$	0.100	0.100	(0.000)	0.100	(0.000)	0.100	(0.000)
$\mu_{Type\,3}$	8.000	8.050	(0.078)	8.025	(0.057)	8.028	(0.076)
$r_{Type\,3}$	0.300	0.298	(0.003)	0.299	(0.002)	0.299	(0.003)
$\mu_{Type\,4}$	10.000	10.018	(0.039)	10.005	(0.027)	10.003	(0.037)
$r_{Type\,4}$	0.300	0.300	(0.001)	0.300	(0.001)	0.300	(0.001)
(II) Subscription eq. (B.2)							
β_0	−2.000	−1.989	(0.149)	−2.005	(0.108)	−1.995	(0.145)
β_{Age}	−0.030	−0.030	(0.001)	−0.030	(0.001)	−0.030	(0.001)
β_{Hsize}	0.100	0.100	(0.011)	0.101	(0.009)	0.101	(0.012)
$\beta_{MiddleIncome}$	0.300	0.295	(0.037)	0.302	(0.027)	0.300	(0.038)
$\beta_{HighIncome}$	0.620	0.613	(0.047)	0.617	(0.032)	0.618	(0.044)
σ_u	1.000	0.992	(0.070)	0.999	(0.051)	0.993	(0.072)
$\sigma_{u,\mu}$	0.500	0.504	(0.028)	0.503	(0.021)	0.503	(0.030)

Note: The results are based on 300 replications. The standard deviation is in the parenthesis. Each month in the simulation has 4 weeks (28 days). The average percentage of subscription is 30.6%, and 6.3% of the daily usage is zero. The number of latent types was selected by the BIC criterion.

Table 3: Simulation with Price Variation

		Estimates: $n = 1e3, M = 10$			
Parameters	Truth	Price = 1, 2		Price = 1, 1.5, 2	
(I) Usage eq. (B.1)					
γ_1	1.000	1.004	(0.005)	1.004	(0.005)
γ_{Mon}	−4.961	−4.982	(0.025)	−4.983	(0.027)
γ_{Tue}	−3.549	−3.566	(0.020)	−3.564	(0.023)
γ_{Wed}	−2.716	−2.729	(0.020)	−2.727	(0.021)
γ_{Thu}	−3.390	−3.405	(0.021)	−3.404	(0.022)
γ_{Fri}	−1.613	−1.621	(0.018)	−1.620	(0.018)
γ_{Sat}	2.346	2.355	(0.019)	2.355	(0.020)
$\mu_{Type\,1}$	8.000	8.055	(0.087)	8.064	(0.081)
$r_{Type\,1}$	0.100	0.099	(0.001)	0.099	(0.001)
$\mu_{Type\,2}$	10.000	10.014	(0.044)	10.010	(0.043)
$r_{Type\,2}$	0.100	0.100	(0.000)	0.100	(0.000)
$\mu_{Type\,3}$	8.000	8.057	(0.081)	8.053	(0.093)
$r_{Type\,3}$	0.300	0.298	(0.003)	0.298	(0.003)
$\mu_{Type\,4}$	10.000	10.016	(0.039)	10.011	(0.040)
$r_{Type\,4}$	0.300	0.300	(0.001)	0.300	(0.001)
(II) Subscription eq. (B.3)					
β_0	−2.000	−2.005	(0.100)	−1.996	(0.112)
β_{Age}	−0.030	−0.030	(0.002)	−0.030	(0.001)
β_{Hsize}	0.100	0.101	(0.012)	0.100	(0.013)
$\beta_{MiddleIncome}$	0.300	0.299	(0.040)	0.298	(0.042)
$\beta_{HighIncome}$	0.620	0.617	(0.050)	0.612	(0.052)
δ_{Tenure}	−0.200	−0.199	(0.034)	−0.201	(0.038)
σ_u	1.000	0.998	(0.057)	0.994	(0.066)
$\sigma_{u,\mu}$	0.500	0.506	(0.031)	0.507	(0.028)
$\sigma_{u,Tenure}$	0.500	0.493	(0.053)	0.495	(0.049)

Note: The results are based on 300 replications. The standard deviation is in the parenthesis. Each month in the simulation has 4 weeks (28 days). The average percentage of subscription is 30%, and 5.6% of the daily usage is zero. The number of latent types was selected by the BIC criterion.

B.2 Results

Table 2 reports the simulation results without price variation. We consider three experiments with varying sample size: $(n = 1000, M = 10)$, $(n = 2000, M = 10)$, $(n = 2000, M = 5)$. In all experiments, our estimator work very well with negligible bias that is likely due to finite sample. Comparing the experiments with $n = 1000, M = 10$ and $n = 2000, M = 10$, we observe the decreasing of standard deviation indicating the consistency of our estimator. Comparing the experiments with $n = 1000, M = 10$ and $n = 2000, M = 5$ (these two panel data have the same sample size), we do not observe significant difference in the estimation performance. This suggests that long panel data (big M) are not necessary—one can increase the cross section sample size (big n) to achieve similar estimation accuracy.

Table 3 reports the simulation results with price variation using the specification in eq. (B.3). In this case, we fix the sample size at $n = 1000, M = 10$, but change the number of prices. The first experiment has two prices, and the second has three prices. First, our estimator perform well in both cases with price variation. Second, it seems that an additional price does not affect the estimation performance. We do not observe significant change of the standard deviation when there are three prices as opposed to two prices. This is fortunate in application, in which the price variation is usually limited.

References

Lewbel, Arthur. 2014. "An Overview of the Special Regressor Method." In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and-Statistics*, Jeffrey S. Racine, Liangjun Su and Aman Ullah, Chapter 2, 38–64. Oxford University Press.