# Can Willingness to Pay be Identified without Price Variation?

## What Big Data on Usage Tracking Can (and Cannot) Tell Us

Cheng Chou [1] Vineet Kumar [2]

[1]University of Leicester, UK

[2]Yale University, USA

UCL, September 24 2019

| Industry | Product or Service | Price ($) | Period | Total subscribers |
|---|---|---|---|---|
| *Media & Entertainment* | Netflix | 9.99 | 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 | – |

# Willingness to Pay (WTP) in marketing

- WTP is essential for customer-driven marketing strategy and marketing mix.
  - estimating demand curve;
  - market targeting;
  - value pricing;
  - product and service decisions (quality, features, etc.)

# Current practice in recovering WTP from data

- Identify the distribution of WTP in a market *with price variation*.
  - Source of price variation: segmented pricing, promotional pricing etc.
  - Current method: **revealed preference** (discrete choice model, auction, Guadagni and Little, 1983, Danthurebandara et al., 2011, Train and Weeks, 2005, Lewbel et al., 2011, etc.), and **conjoint analysis** (Green and Rao, 1971, Green and Srinivasan, 1978, Rao, 2014, etc.)
- **Price variation is essential for nearly the entire literature.**

# Identify the distribution of WTP without price variation

- Interested in the distribution of WTP for subscription service
  - **Challenge: absence of price variation.**
  - Opportunity: big usage tracking data.
  - Why not run price experiment? (Ariely, 2010)
- Research questions:
  - *Can Willingness to Pay Distribution be Identified without Price Variation?*
  - Additional value of price variation?

## Key insight from a simple example

- How does a firm recover the WTP distribution without Price variation?
- Usage variation $\Rightarrow$ variation in **price of one unit of usage**

| User | # Gym Visits Per Month | Monthly Fee | "Price" Per Visit |
|---|---|---|---|
| 1 | 10 | £30 | £3 |
| 2 | 5 | £30 | £6 |

- (WTP for Gym Membership) = (WTP per [Average] Visit) $\times$ (Expected # Visits).
- Subscribe if (WTP per Visit) $>$ ("Price" Per Visit).
- Market share is informative for distribution of (WTP per Visit).

- Suppose only two usages 5 and 10 visits in population with equal probability.

| *Users visiting 5 times* | *Users visiting 10 times* |
|---|---|
| "Price" per visit $= £6$ | "Price" per visit $= £3$ |
| Retention rate $= 60\%$ | Retention rate $= 80\%$ |
| P(WTP per visit $< £6) = 0.4$ | P(WTP per visit $< £3) = 0.2$ |

$$
\begin{aligned}
&P(\text{WTP for Gym Membership} < £30) = \\
&\qquad P(\#\ \text{Visits} = 5) \times P(\text{WTP per visit} < £6) + \\
&\qquad P(\#\ \text{Visits} = 10) \times P(\text{WTP per visit} < £3) = \\
&\qquad\qquad 0.5 \times 0.4 + 0.5 \times 0.2 = 0.3.
\end{aligned}
$$

## Steps in finding the distribution of WTP in general

| | | | |
|---|---|---|---|
| *Step 0.* | Collect data: (a) subscription/churn choices, (b) usage, (c) product attributes, (d) consumers characteristics. | | |
| *Step 1.* | Market share or retention rate for different usage levels | $\Rightarrow$ | Distribution of WTP per unit of usage |
| *Step 2.* | Usage data | $\Rightarrow$ | Distribution of expected usage |
| *Step 3.* | Combine distribution of WTP per visit and usage | $\Rightarrow$ | Distribution of WTP for service |
| *Step 4.* | Estimate elasticities, demand curve, and do counterfactual | | |

## Make inference about the WTP distribution in the entire population

### Why is it challenging?

- Heterogeneity
  - WTP per visit
  - usage (expected # of visits)
- Correlation between WTP per visit and usage
- Unobserved usage by nonsubscribers
- Selection: nonsubscribers are different from subscribers in unobservable ways

Our approach accommodates all the above issues.

# Contribution

- Main contribution: a novel method to identify & estimate semiparametrically the conditional distribution of WTP given customer characteristics and product features **when only usage variation is present**.
- No research that demonstrates how to obtain the WTP distribution in the **absence of price variation**.
- Closest research: Nevo et al. (2016).

# Big usage data of YBOX, a music streaming service

- YBOX is a music streaming service targeting Southeast Asia. 80% market share, $> 10$ million users.
- 1 million users data (Jan 2015–Feb 2017):
  - subscription history
  - daily # of seconds listening music with the service
  - basic demographics (age and gender)
- **Three prices** (*roughly* \$5 [56.8%], \$4 [17.6%], \$3 [23.9%]) for the same monthly music streaming service
- Average monthly listening hours range from less than 1 hour to more than 150 hours.

# Descriptive statistics of YBOX

Table: Descriptive Stat YBOX Users

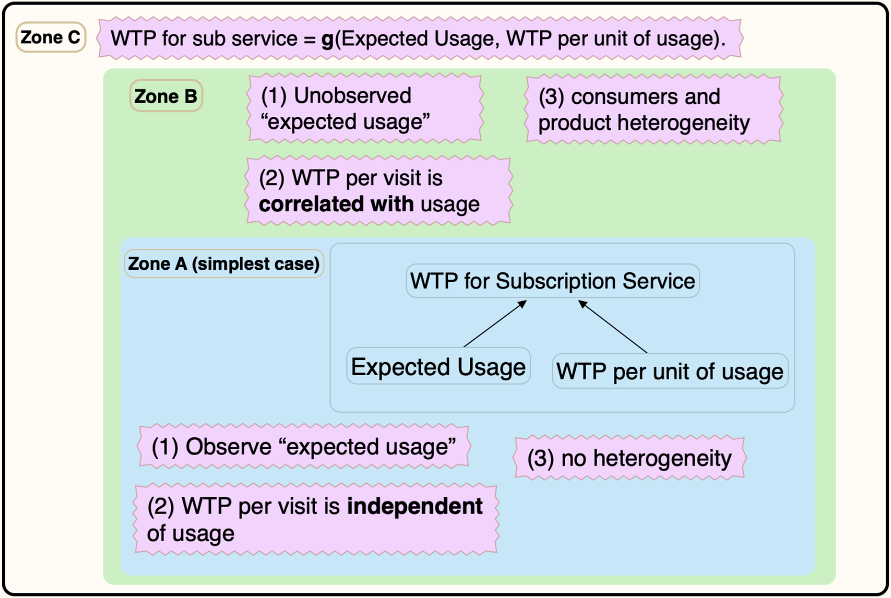
| | Users churned | Users not churn |
|---|---|---|
| Average daily listening hours | 1.17 | 1.33 |
| | (0.99) | (1.09) |
| Age | 29.53 | 29.92 |
| | (8.25) | (8.14) |
| Is male | 0.51 | 0.55 |
| *n* of obs | 1211 | 7487 |

- Randomly sampled 8698 users from one city.

# Result preview: estimated WTP for YBOX monthly plan

| | Estimated Mean of Log WTP | |
|---|---|---|
| Age group | Men | Women |
| Before college | log($16) | log($16) |
| In college | log($24) | log($21) |
| 23–30 | log($33) | log($30) |
| $>30$ | log($49) | log($56) |

# Outline


Zone C
WTP for sub service = **g**(Expected Usage, WTP per unit of usage).
Zone B
(1) Unobserved "expected usage"
(3) consumers and product heterogeneity
(2) WTP per visit is **correlated with** usage
Zone A (simplest case)
WTP for Subscription Service
Expected Usage
WTP per unit of usage
(1) Observe "expected usage"
(3) no heterogeneity
(2) WTP per visit is **independent** of usage

# Model setup: simplest case (ZONE A)

## Notation:

- $i$ indicates a consumer
- WTP for **service**: $W_i$
- WTP **per unit of usage**: $\alpha_i$
- Expected usage: $Q_i^*$
- Decision: $S_i = 1$ (sub) and $= 0$ (not)

## Assumptions in ZONE A

1. $W_i = \alpha_i Q_i^* \Rightarrow \ln W_i = \ln \alpha_i + \ln Q_i^*$.
2. $Q_i^*$ is observed for both subscribers and nonsubscribers.
3. $\alpha_i \perp\!\!\!\perp Q_i^*$.

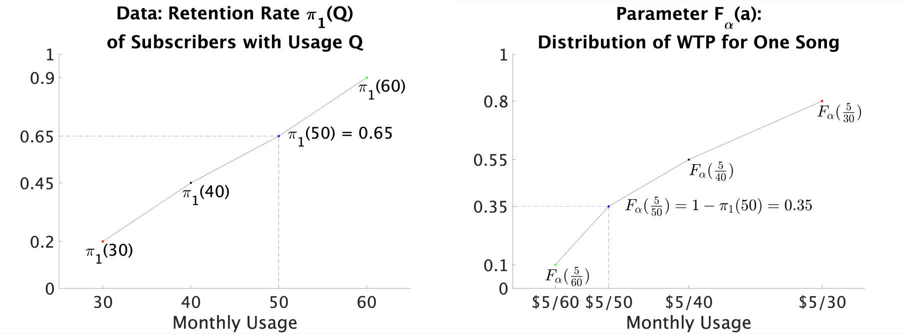
Microfoundation of usage

Decision rule: $S_i = \mathbb{I}(P < W_i) = \mathbb{I}(P < \alpha_i Q_i^*)$.

# Nonparametric estimation of $F_\alpha(\alpha_i)$ (**Step 1**)

Decision rule $S_i = \mathbb{I}(P < \alpha_i Q_i^*)$ implies that

$$F_\alpha(P/Q) \equiv \Pr(\alpha_i \leq P/Q) = 1 - \underbrace{\Pr(S_i = 1 \mid Q_i^* = Q)}_{\pi_1(Q)} \quad \Rightarrow \quad \boxed{F_\alpha(a) = 1 - \pi_1\left(\frac{P}{a}\right)}$$



# WTP: From usage $\Rightarrow$ subscription (**Step 2 and 3**)

- If
  - $F_\alpha(\alpha_i)$ is known,
  - $Q_i^*$ is observed, so $F_Q(Q_i^*)$ is known, (**Step 2**)
  - $\alpha_i \perp\!\!\!\perp Q_i^*$,

  the distribution $F_W(W_i)$ is calculated by using $W_i = \alpha_i Q_i^*$.

$$F_W(w) = \int_0^\infty F_\alpha(w/q)\, \mathrm{d}\, F_{Q^*}(q).$$

- It can shown that

$$\boxed{F_W(w) = 1 - \mathrm{E}\Big(\pi_1\Big(\frac{PQ_i^*}{w}\Big)\Big).}$$

## Algorithm for ZONE A

*Step 1:* estimate $\pi_1(q) \equiv \Pr(S_i = 1 \mid Q_i^* = q)$;

*Step 2a:* estimate $\hat{F}_\alpha(a) = 1 - \hat{\pi}_1(P/a)$.

*Step 2b:* estimate $\hat{F}_W(w) = 1 - n^{-1} \sum_{i=1}^{n} \hat{\pi}_1(PQ_i^*/w)$.

# Applications and counterfactual analysis (**Step 4**)

- Demand function: $D(p) \equiv \Pr(S_i = 1 \mid P = p) = 1 - F_W(p)$.
- Pricing a shorter subscription plan based on **higher frequency usage** data.
  - Let $Q^*_{i,wk}$ be the expected usage in *week* $wk = 1, \ldots, 4$.

$$W_{i,wk} = \alpha_i Q^*_{i,wk}.$$

  - Note that in general $Q^*_{i,wk}$ and $Q^*_i/4$ ($Q^*_i$ is monthly usage) have different distributions.
- More possible counterfactual analysis in ZONE B.

Why not just price a weekly plan by dividing the monthly price by 4?

| | | |
|---|---|---|
| $W_{i,wk}$ | | ${}^{W_{i,month}}\!/_{4}$ |
| $W_{i,wk} = \alpha_i Q^*_{i,wk}$ | | ${}^{W_{i,month}}\!/_{4} = \alpha_i {}^{Q^*_{i,month}}\!/_{4}$ |
| Distribution of $Q^*_{i,wk}$ | $\stackrel{?}{=}$ | Distribution of ${}^{Q^*_{i,month}}\!/_{4}$ |

# Summary by R practice

- https://chengchou.shinyapps.io/WTPV1/

# Objectives in ZONE B

- Accommodate consumers and product **heterogeneity**.
- Deal with **unobserved** "expected usage" (especially by non-subscribers).
- Allow the **correlation** between WTP per unit of usage ($\alpha_i$) and expected usage ($Q_i^*$).
- Selection issue: subscribers and nonsubscribers are different in unobservable ways.

# Parameters of interest in ZONE B

- $X_i$: observed consumer and product characteristics
- $F_W(W_i \mid X_i) \Rightarrow$
  - Demand functions in different market segments

$$\begin{aligned} D(p \mid Male_i = 1) &= \Pr(S_i = 1 \mid P = p, Male_i = 1) \\ &= 1 - F_W(p \mid Male_i = 1). \end{aligned}$$

  - Revenue (percent) change w.r.t. the change of product features.

$$\frac{D(P \mid StreamHD_i = 1) - D(P \mid StreamHD_i = 0)}{D(P \mid StreamHD_i = 0)}$$

- $\mathrm{E}(\ln W_i \mid X_i) \Rightarrow$ elasticities

$$\mathrm{E}\left(\frac{\partial \mathrm{E}(\ln W_i \mid X_i)}{\partial X_{ij}}\right), \text{ e.g., } X_{ij} \text{ is ad. spending}$$

# Two Dimensional Heterogeneity

| | *WTP per unit of usage* | *Expected Usage* |
|---|---|---|
| *Model* | $\ln \alpha_{it} = \beta' X_{1it} + U_{it}$ | $\ln Q^*_{it} = \gamma' X_{2it} + V_i$ for all<br>$\ln Q_{it} = \ln Q^*_{it} + \varepsilon_{it}$ for subrs |

*Remarks*

- Examples of $X_{1it}$ and $X_{2it}$ in Netflix.

  $X_{1it}$: age, gender, size/quality of movie library

  $X_{2it}$: age, gender, local bandwidth, weather
- $U_{it}$ and $V_i$ could be correlated.
- Denote $\eta_{it} = U_{it} + V_i$
- $W_{it} = \alpha_{it} Q^*_{it}$ implies

$$\begin{aligned} \ln W_{it} &= \ln \alpha_{it} + \ln Q^*_{it} \\ &= \beta' X_{1it} + \gamma' X_{2it} + \eta_{it}, \end{aligned}$$

- No parametric assumption for $U_{it}$ or $V_i$.

# Two Dimensional Heterogeneity

| | *WTP per unit of usage* | *Expected Usage* |
|---|---|---|
| *Model* | $\ln \alpha_{it} = \beta' X_{1it} + U_{it}$ | $\ln Q^*_{it} = \gamma' X_{2it} + V_i$ for all<br>$\ln Q_{it} = \ln Q^*_{it} + \varepsilon_{it}$ for subrs |
| *Algorithm* | *Step 1:* Fixed effect reg $\Rightarrow \gamma$ and $\tilde{Q}_{it} = \gamma' X_{2it}$. ▸ Time invariant $X_{2it}$ | |
| *for Zone B* | *Step 2:* Est $\beta$ and $\mathrm{E}(\ln W_{it} \mid X_{it})$ using Theorem 1 (next slide). | |
| | *Step 3:* Est $\pi_2(X_{1it}, \tilde{Q}_{it}) = \mathrm{E}(S_{it} \mid X_{1it}, \tilde{Q}_{it})$ nonparam. | |
| | *Step 4a:* Est $F_\eta(\eta \mid X_{1it}) = 1 - \pi_2(X_{1it}, \ln P - \beta' X_{1it} - \eta)$ | |
| | *Step 4b:* Est $F_W(w \mid X_{it}) = 1 - \pi_2(X_{1it}, \ln P - \ln w + \gamma' X_{2it})$ | |

## Theorem 1

*Under big support (next slide) and other conditions,*

1. $\beta$ *can be estimated by running OLS of*

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

   *on* $X_{1it}$.

2. $\mathrm{E}(\ln W_{it} \mid X_{it}) = \beta' X_{1it} + \gamma' X_{2it}$.

## Key Assumption

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

- Support condition implies **excluded variable requirement:** need variables that affect usage but not the WTP per unit of usage, e.g. weather. Theorem with endogenous $X_{1it}$

# Benefits of having price variation

## What big data on usage cannot tell us? ZONE C

- So far $W_i = \alpha_i Q_i^* \iff$

$$\begin{aligned} W_i &= \exp(\ln \alpha_i + \ln Q_i^*) \\ &= \exp(\beta' X_{i1} + U_i + \ln Q_i^*). \end{aligned}$$

- We want to relax multiplicative form by letting

$$W_i = g(\ln \alpha_i + \ln Q_i^*) = g(\beta' X_{i1} + U_i + \ln Q_i^*),$$

  where $g(\cdot)$ is unknown.
- Can we identify $F_W(W_i)$ and $g(\cdot)$ in the absence of price variation? No. Two unknowns $F_U$ and $g$.
- **Price variation** can identify $g(\cdot;\delta)$, where $\delta$ is a vector of unknown parameters, thus relax multiplicative form.

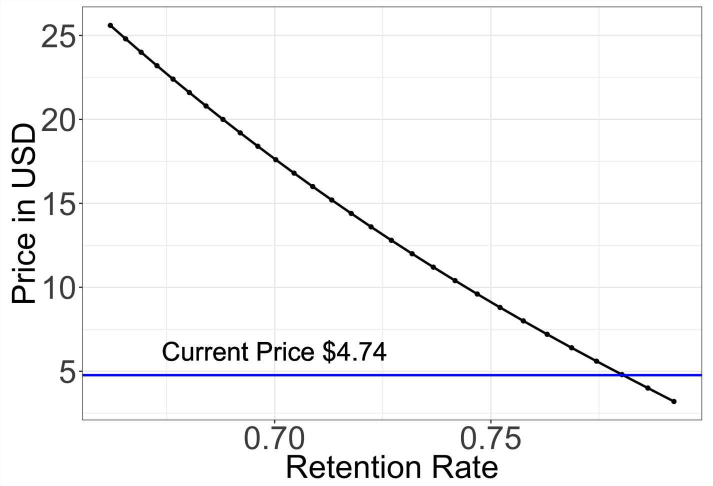
- Age elasticity: 0.04 (0.01).
- Men are willing to pay 26.23% more than women.
- For Age $>$ 30, the mean monthly listening hours of men and women are 19.6 and 21.5 hours.

| | Ested Mean of Log WTP | |
|---|---|---|
| Age | Men | Women |
| Before college | log($16) | log($16) |
| In college | log($24) | log($21) |
| 23–30 | log($33) | log($30) |
| $>$ 30 | log($49) | log($56) |

*WTP and Usage Elasticities (Hundredths) Estimates*

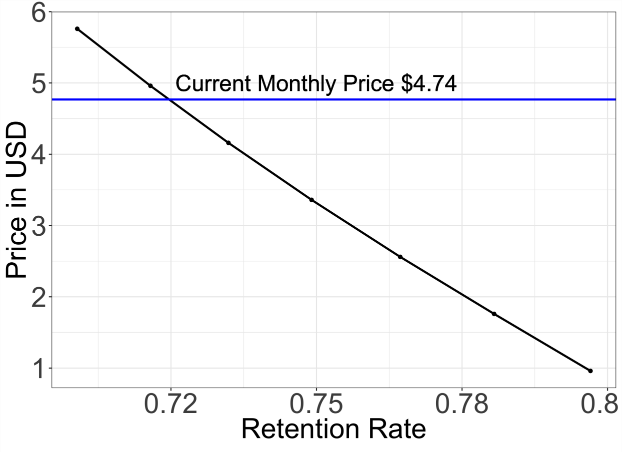
| | WTP | Usage (All Users) |
|---|---|---|
| Age | 3.98 | $-2.00$ |
| | (1.19) | (0.22) |

# Estimated demand curve for current **monthly** streaming plan




**Inelastic.** Is increasing the current price profitable? Depends on marginal cost, but probably YES.

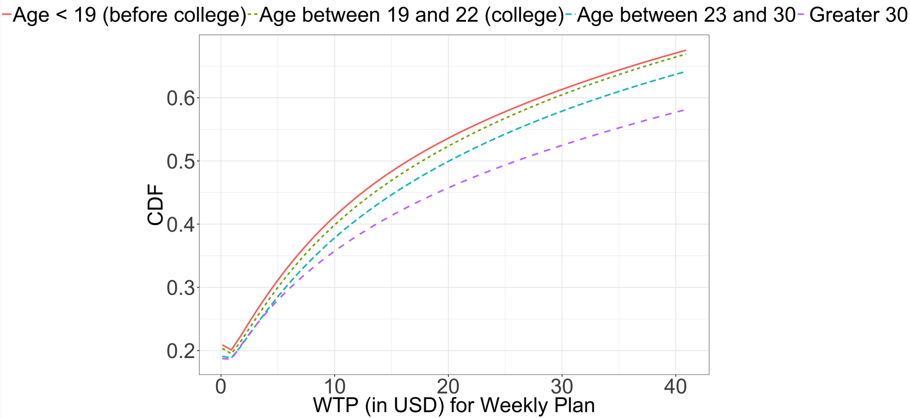
# Demand for **counterfactual weekly** streaming plan



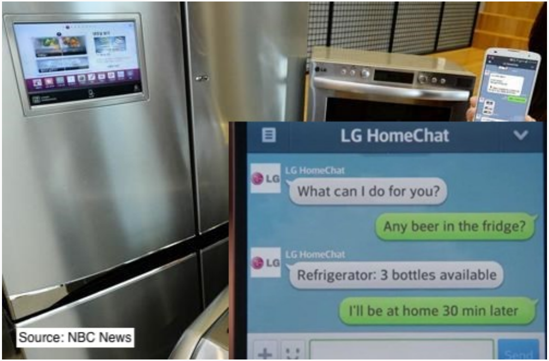

To obtain 75% retention rate,

- the monthly price is \$8.8,
- the weekly price is \$3.2 HIGHER than \$2.2 = \$8.8 / 4.

# Estimated distribution of the WTP for counterfactual weekly plan



# Conclusion: a bigger picture (of a fridge)




- Essentially, we need the separation of purchase (subscription) and consumption (usage).
- Such separation also holds in packaged goods (beer)—but we did not track the usage.
- 5G and Internet of Things could enable the tracking.

# BACKUP SLIDES

# Include time invariant $X_{2it}$

- We know how to estimate $\beta$ from

$$S_{it} = \mathbb{I}(\beta' X_{1it} + \tilde{Q}_{it} - \ln P + \eta_{it} > 0).$$

- Suppose $X'_{2it} = (X'_{2i,a}, X'_{2it,b})$. We still write

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

We cannot identify $\gamma_a$, but can identify $\gamma_b$ from FE.

- Define $\tilde{Q}_{it} = \gamma'_b X_{2it,b}$, and we have

$$S_{it} = \mathbb{I}(\beta' X_{1it} + \gamma'_a X_{2i,a} + \tilde{Q}_{it} - \ln P + \eta_{it} > 0).$$

When $X_{1it}$ and $X_{2i,a}$ are different, we can separately identify $\beta$ and $\gamma_a$.

Back to 'Two Dim Heterogeneity'

# Usage Model: Microfoundations

- Earlier with exogenous $Q_i^*$ - the usage model was not specified.
- Each subscription setting might have a different micro-model
- Spotify Example: Consumers solve time allocation problem between listening to music and the outside option.
  - $u(q, \nu, T; \theta) = \theta_1 q \underbrace{\nu}_{\text{shock}} + \theta_2 q^2 + (\underbrace{T}_{\text{Time budget}} - q)$
  - Obtain the usage decision rule: $q(\nu, T; \theta)$
  - $Q_i^* = E_\nu \left[ q(\nu, T; \theta) \right]$
- Issue to consider: Does $\theta$ (parameters of micromodel) depend on the counterfactual?
  - If it does not: we can use the same micromodel in counterfactuals. Don't really need to estimate $\theta$.
  - if it does: we need to estimate the micromodel parameters $\theta$

◂ Back to simplest model

Back to assumptions

## Theorem 2

*Under certain conditions,*

1. $\beta$ *can be estimated by 2SLS of*

$$Y_{2it} = \frac{S_{it} - \mathbb{I}(\tilde{Q}_{it} - \ln P > 0)}{f_{\tilde{Q}}(\tilde{Q}_{it} \mid X_{1it}, Z_{it})}.$$

*on* $X_{1it}$ *with IV* $Z_{it}$.

2. $\mathrm{E}(\ln W_{it} \mid X_{1it}, X_{2it}) = \beta' X_{1it} + \gamma' X_{2it} + \mathrm{E}(\eta_{it} \mid X_{1it}, X_{2it})$, *where* $\mathrm{E}(\eta_{it} \mid X_{1it}, X_{2it}) = \mathrm{E}(H_{2it} \mid X_{1it}, X_{2it})$ *with*

$$H_{2it} = \mathrm{E}\left( \frac{S_{it} - \mathbb{I}(\ln P - \beta' X_{1it} - \tilde{Q}_{it} \leq 0)}{f_{\tilde{Q}}(\tilde{Q}_{it} \mid X_{1it}, Z_{it})} \,\middle|\, X_{1it}, Z_{it} \right).$$

Why are companies reluctant to experiment with price?

- Fairness (both cross-sectional and dynamic)
- Brand image impact

ORGANIZATIONAL CULTURE

# Column: Why Businesses Don't Experiment

by Dan Ariely

FROM THE APRIL 2010 ISSUE

Save Share 0 Comment Text Size Print $6 Buy Copies

A few years ago, a marketing team from a major consumer goods company came to my lab eager to test some new pricing mechanisms using principles of behavioral economics. We decided to start by testing the allure of "free," a subject my students and I had been studying. I was excited: The company would gain insights into its customers' decision making, and we'd get useful data for our academic work. The team agreed to create multiple websites with different offers and pricing and then observe how each worked out in terms of appeal, orders, and revenue.

Several months later, right before we were due to go live, we had a meeting about the final details of the experiment—this time with a bigger entourage from marketing. One of the new members noted that because we were extending differing offers, some customers might buy a product that was not ideal for them, spend too much money, or get a worse deal overall than others. He was correct, of course. In any experiment, someone gets the short end of the stick. Take clinical medical trials, I said to the team. When testing chemotherapy treatments, some patients suffer more so that, down the road, others might suffer less. I hoped this put it in perspective. Fortunately, I said, price testing household products requires far less suffering than chemo trials.

But I could tell I was losing them. In a sense, I was impressed. It was a beautiful human sentiment they were conveying: We care about all customers and don't want to treat any one of them unfairly. A debate ensued among the group: Are we willing to sacrifice some customers "just" to learn how the new pricing approaches work?

They hedged. They asked me what *I* thought the best approach was. I told them that I was willing to share my intuition but that intuition is a remarkably bad thing to rely on. Only an experiment gives you the evidence you need. In the end, it wasn't enough to convince them, and they called off the project.

This is a typical case, I've found. I've often tried to help companies do experiments, and usually I fail spectacularly. I remember one company that was having trouble getting its bonuses right. I suggested they do some experiments, or at least a survey. The HR staff said no, it was a miserable time in the company. Everyone was unhappy, and management didn't want to add to the trouble by messing with people's bonuses merely for the sake of learning. *But the employees are already unhappy*, I thought, and the experiments would have provided evidence for how to make them less so in the years to come. How is that a bad idea?

Back to business

# Identify the valuation of the outside option

- Let $g^*(c;\delta) = e^c - \delta$ with unknown $\delta$. We have

$$\begin{aligned} S_i &= \mathbb{I}(g^*(\ln \alpha_i + \ln Q_i^*; \delta) > P) \\ &= \mathbb{I}(\exp(\ln \alpha_i + \ln Q_i^*) - \delta > P) \\ &= \mathbb{I}(\alpha_i Q_i^* - P > \delta). \end{aligned}$$

- We have been letting $\delta = 0$.
- Two prices $P_A$ and $P_B$ for the same service $\Rightarrow$ identification of $\delta$.

$$
\mathrm{E}\left(\frac{S_{i,A}}{f_{\tilde{Q}_A}(\tilde{Q}_{i,A})} - \frac{S_{i,B}}{f_{\tilde{Q}_B}(\tilde{Q}_{i,B})}\right) =
$$

$$
\mathrm{E}\left(\frac{\mathbb{I}(\tilde{Q}_{i,A} - \ln(P_A + \delta) > 0)}{f_{\tilde{Q}_A}(\tilde{Q}_{i,A})} - \frac{\mathbb{I}(\tilde{Q}_{i,B} - \ln(P_B + \delta) > 0)}{f_{\tilde{Q}_B}(\tilde{Q}_{i,B})}\right),
$$

.

- The only unknown is $\delta$ in the above moment equation.
- To see more intuition, it can be shown that the above is identical to

$$
\int_{\mathcal{Q}} \int_{\eta} (S_{i,A} - S_{i,B}) \, \mathrm{d}\, F_{\eta}(\eta) \, \mathrm{d}\, q = \ln\left(\frac{P_B + \delta}{P_A + \delta}\right).
$$