Identifying Search and Stockpiling Behavior: Experimental and Field Data

Andrew T. Ching

Johns Hopkins University

Dec 12, 2025

DSE Winter School Lecture, HKU

Plan

- Two papers:
- Ching, A.T. and M. Osborne (2020) “Identification and Estimation of Forward-looking Behavior: The Case of Stockpiling,” *Marketing Science*.
- Osgouei A.J., A.T. Ching, S.S. Tehrani, B. Ratchford (2025) “Estimating Position and Social Influence Effects in Online Search,” forthcoming in *Marketing Science*.
- Ching and Osborne (2020) uses field data.
- Osgouei et al. (2025) uses experimental data.

Identification and Estimation of Forward-looking Behavior: The Case of Stockpiling

Andrew T. Ching and Matthew Osborne

Motivations

- Discount factor is often set according to the prevailing interest rate -- the rational expectation assumption.
- Researchers often use the reason: Without any model restrictions (non-parametric), dynamic discrete choice models are not identified (Rust, 1994).
- Here, we propose a specific exclusion restriction motivated by institutional setting to identify the discount factor, β.
- The context is consumer stockpiling.
- Our approach is closely related to Fang and Wang (2015) and Abbring and Daljord (2020) for identification of discount factor.
- Extend pioneer works in stockpiling models: Erdem, Imai and Keane (2003), Hendel and Nevo (2006).

Exclusion Restrictions

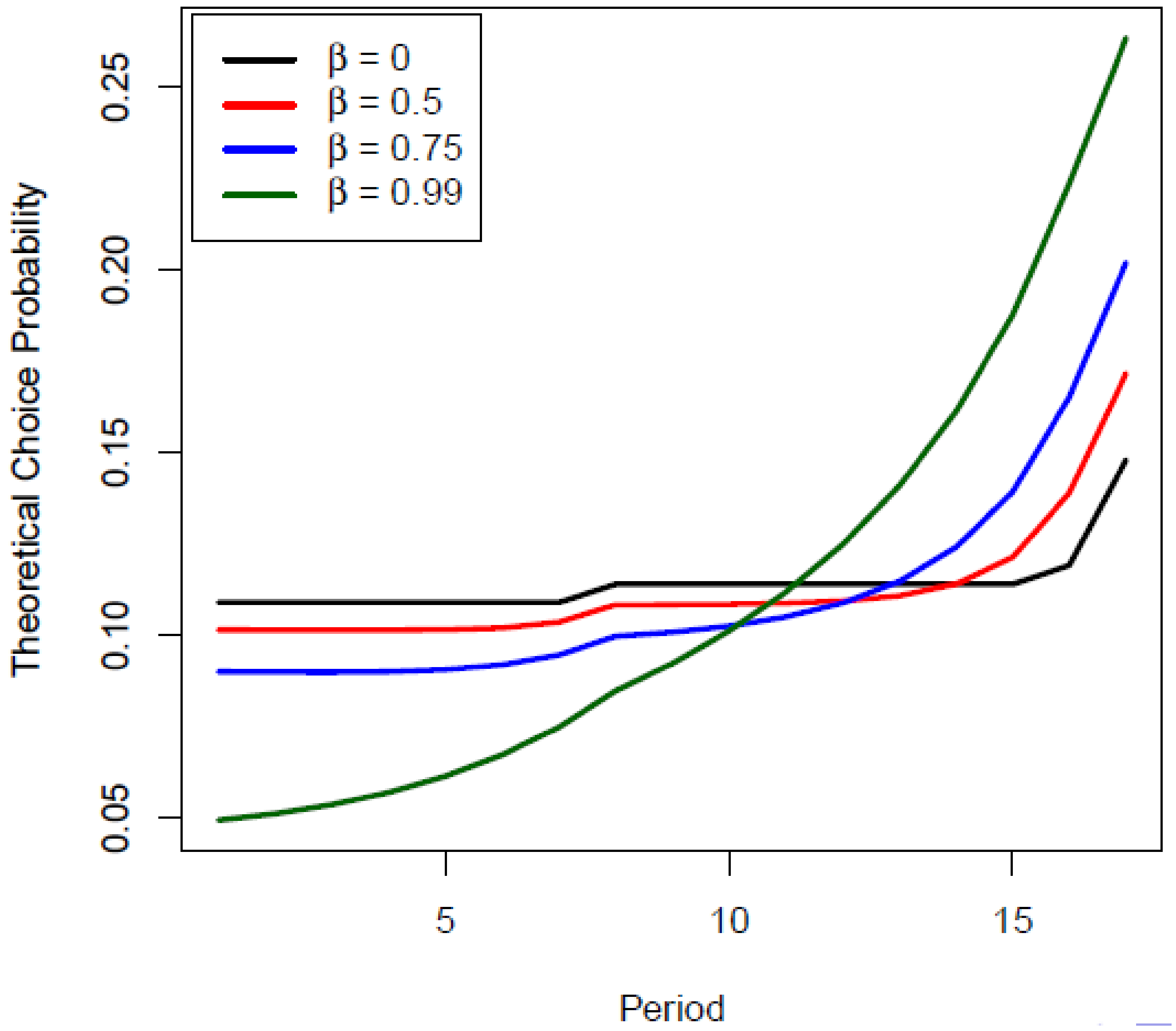
- Defn: There exists two values of a state variable, where for some choice alternatives, the current flow utilities remain unchanged, but the expected future value could differ depending on β.
- The inventory variable can provide exclusion restrictions.
- Why? In consumer package goods, the storage cost of inventory is often flat.
  - It may be a function of the number of packages (e.g., laundry detergent) – this can lead to a step function.
  - A consumer may set aside some space in the laundry room to store laundry detergent. The marginal cost of storing an extra bottle is zero until hitting that space constraint.

How Inventory help identify β

- Consider **price is fixed**.
- Consider a consumer has only one bottle of detergent left.
- As the amount of the bottle decreases, the possibility of having stockout increases in the future, while the storage cost remains unchanged.
  - The bottle remains and takes up the same amount of space.
- A forward-looking consumer’s purchase probability should increase as inventory drops (<u>even though price is fixed</u>).
- But a myopic consumer’s purchase probability will remain constant even when inventory drops – it will only increase when inventory runs out.

Identification with Observed Inventory (intuition)

- Set I = 16 in period 9, and assume inventory decrease by 1 unit every period.
- Plot the choice prob.



Model Assumptions

- Single product available in one package size.
- Package size is $b = 8$ units.
- Consumption need is fixed at $c = 1$ (we can relax this).
- Purchase price is $p > 0$ (fixed over time for now).
- Price coefficient: α.
- Stockout cost is ν.
- Discount factor is β.
- State is consumer inventory level I.
- Consumer decision is to buy or not buy a single package.
- Maximum storage capacity is $M = 3$ packages.
- Consumption utility γ.
- Set $\gamma = 0$ for simplicity.

Proposition: β is identified

$$\begin{aligned} v_1(I) &= -\alpha p - \omega_2 + \beta V(I+b-1), \quad (15) \\ v_0(I) &= -\omega_1 + \beta V(I-1). \quad (16) \end{aligned}$$

We are interested in considering the difference in choice-specific value for these two options because they are simply a function choice probabilities by Hotz and Miller (1993)'s Inversion Theorem.

$$\begin{aligned} v_1(I) - v_0(I) &= -\alpha p - (\omega_2 - \omega_1) + \beta[V(I+b-1) - V(I-1)] \quad (17) \\ \Delta v(I) &= -\alpha p - \Delta\omega(2,1) + \beta[V(I+b-1) - V(I-1)], \quad (18) \end{aligned}$$

where $\Delta v(I) \equiv v_1(I) - v_0(I)$; $\Delta\omega(2,1) \equiv \omega_2 - \omega_1$.

Note that our exclusion restriction assumption implies that if we take the difference of $\Delta v(.)$ at two different values of I that satisfy the exclusion restriction requirements, we can difference out the current utility components.

$$\begin{aligned} \Delta v(I+1) - \Delta v(I) &= \beta[V(I+b) - V(I) - (V(I+b-1) - V(I-1))] \\ &= \beta[V(I+b) - V(I+b-1) - (V(I) - V(I-1))] \quad (19) \end{aligned}$$

The following argument is new (slides# 8-10), collaborated with Sihan Zhai

$$
\begin{aligned}
&\Delta \log(\hat{P}(l+1)) - \Delta \log(\hat{P}(l)) \\
=&\beta(V(l+b) - V(l) - V(l+b-1) + V(l-1)) \\
=&\beta(v_0(l+b) - v_0(l) - v_0(l+b-1) + v_0(l-1) \\
&- [\log(1-P(l+b)) - \log(1-P(l)) - \log(1-P(l+b-1)) + \log(1-P(l-1))])
\end{aligned}
$$

Let $\hat{M} \equiv \log(1-\hat{P}(l+b)) - \log(1-\hat{P}(l)) - \log(1-\hat{P}(l+b-1)) + \log(1-\hat{P}(l-1))$, then

$$
\begin{aligned}
&\Delta \log(\hat{P}(l+1)) - \Delta \log(\hat{P}(l)) \\
=&\beta(v_0(l+b) - v_0(l) - v_0(l+b-1) + v_0(l-1) - \hat{M}) \\
=&\beta(v_1(l) - v_0(l) - v_1(l-1) + v_0(l-1) - \hat{M}) \\
=&\beta(\Delta \log(\hat{P}(l)) - \Delta \log(\hat{P}(l-1)) - \hat{M})
\end{aligned}
$$

Recall from Arcidiacono and Miller (2011)

$$V(x) = \log(\exp(v_0(x)) + \exp(v_1(x))) + \gamma$$

$$1 - P(x) = \frac{\exp(v_0(x))}{\exp(v_0(x)) + \exp(v_1(x))}$$

$$\log(1 - P(x)) = v_0(x) - \log(\exp(v_0(x)) + \exp(v_1(x)))$$

Therefore,

$$V(x) = v_0(x) - \log(1 - P(x)) + \gamma$$

We know

$$\begin{aligned} v_0(I+b) &= -\omega_2 + \beta V(I+b-1) \\ v_1(I) &= -\alpha p - \omega_2 + \beta V(I+b-1) \end{aligned}$$

Therefore,

$$v_0(I+b) = v_1(I) + \alpha p$$

Similarly,

$$\begin{aligned} v_0(I+b-1) &= -\omega_2 + \beta V(I+b-2) \\ v_1(I-1) &= -\alpha p - \omega_2 + \beta V(I+b-2) \end{aligned}$$

Therefore,

$$v_0(I+b-1) = v_1(I-1) + \alpha p$$

Below is the original proof from Ching and Osborne (2020) requires two exclusion restrictions, following slide #7

$$v_0(I) = V(I) - \ln\left(1 + \frac{P_1(I)}{P_0(I)}\right). \tag{20}$$

Suppose WLOG that an individual has a single package in inventory. Then we can rewrite Eq(16) as,

$$\begin{aligned} V(I-1) &= \frac{1}{\beta}\left(v_0(I) + \omega_1\right) \\ &= \frac{1}{\beta}\left(V(I) - \ln\left(1 + \frac{P_1(I)}{P_0(I)}\right) + \omega_1\right), \end{aligned} \tag{21}$$

$$
\begin{aligned}
V(I) - V(I-1) &= V(I) - \frac{1}{\beta}\left(V(I) - \ln\left(1 + \frac{P_1(I)}{P_0(I)}\right) - \omega_1\right) \\
&= \left(1 - \frac{1}{\beta}\right) V(I) + \frac{1}{\beta}\left(\ln\left(1 + \frac{P_1(I)}{P_0(I)}\right) + \omega_1\right). \qquad (22)
\end{aligned}
$$

Moreover,

$$
V(I+b) - V(I+b-1) = \left(1 - \frac{1}{\beta}\right) V(I+b) + \frac{1}{\beta}\left(\ln\left(1 + \frac{P_1(I+b)}{P_0(I+b)}\right) + \omega_2\right). \qquad (23)
$$

Note that because I corresponds to the first package, $I+b$ corresponds to the second package. Therefore, it is ω_2 (storage cost for two package) that appears in Eq(23).

Then using Eq(22) and (23) for $V(I) - V(I-1)$, we can rewrite Eq(19) as

$$
\begin{aligned}
\Delta v(I+1) - \Delta v(I) &= \beta\left[\left(\left(1 - \frac{1}{\beta}\right) V(I+b) + \frac{1}{\beta}\left(\ln\left(1 + \frac{P_1(I+b)}{P_0(I+b)}\right) + \omega_2\right) - \right.\right. \\
&\quad \left.\left.\left(\left(1 - \frac{1}{\beta}\right) V(I) + \frac{1}{\beta}\left(\ln\left(1 + \frac{P_1(I)}{P_0(I)}\right) + \omega_1\right)\right)\right)\right] \\
&= (\beta - 1)(V(I+b) - V(I)) + \Phi(I) + \Delta\omega(2,1), \qquad (24)
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi(I) &= \ln\left(1 + \frac{P_1(I+b)}{P_0(I+b)}\right) - \ln\left(1 + \frac{P_1(I)}{P_0(I)}\right), \\
\Delta\omega(2,1) &= \omega_2 - \omega_1.
\end{aligned}
$$

We can simplify the above equation to,

$$\Delta v(I) = \frac{1}{\beta}\Delta v(I+1) - \Phi(I) - \frac{2\beta - 1}{\beta}\Delta\omega(2,1) - \frac{\beta - 1}{\beta}\alpha p. \tag{25}$$

To get rid of the last two terms on RHS, we can take advantage of the exclusion restriction again and look at $\Delta v(I+1) - \Delta v(I)$ for values of I where only a single package is held. By applying Eq(25) to $(I+1, I+2)$, we get:

$$\Delta v(I+1) = \frac{1}{\beta}\Delta v(I+2) - \Phi(I+1) - \frac{2\beta - 1}{\beta}\Delta\omega(2,1) - \frac{\beta - 1}{\beta}\alpha p. \tag{26}$$

Now by combining Eq(25) and Eq(26), we have:

$$\Delta v(I+1) - \Delta v(I) \;\; = \;\; \frac{1}{\beta}(\Delta v(I+2) - \Delta v(I+1)) - (\Phi(I+1) - \Phi(I)),$$

and solving for β, we get:

$$\begin{aligned}\beta &= \frac{\Delta v(I+2) - \Delta v(I+1)}{\Delta v(I+1) - \Delta v(I) + \Phi(I+1) - \Phi(I)} \\ &= \frac{\Delta \log(P(I+2)) - \Delta \log(P(I+1))}{\Delta \log(P(I+1)) - \Delta \log(P(I)) + \Phi(I+1) - \Phi(I)},\end{aligned} \tag{27}$$

where

$$\Delta \log(P(I)) \equiv \log(P_1(I)) - \log(P_0(I)) = \Delta v(I).$$

Remark 1

- The proof in Ching and Osborne (2020) requires two exclusion restrictions: (I, I+1) and (I+1, I+2) (Slide #7, 11-14).

- The new proof (Slide #8-10, collaborated with Sihan Zhai): Take advantage of v_0(I+b) - v_1(I), as they both have the same end of period inventory, allowing us to difference out the expected future component. This is inspired by Kong, Dube and Alwan (2025).

- Both proofs lead to the same moment condition that identifies β.

- The rest of the structural parameters can be identified in a straightforward way.

Remark 2

- It is important to highlight that even with fixed price, a stockpiling model with inventory as a state variable is sufficient to identify the discount factor.
  - Kong, Dube and Alwan (2025) misunderstood that price follows Markov is necessary for a stockpiling model.
- Price fluctuation is **<u>not</u>** needed for a stockpiling model to generate dynamic forward-looking decisions
- Forward-looking consumers still have stronger incentive to buy when getting closer to use up inventory.
  - To avoid stock-out cost or no consumption as it is always possible to draw a very negative ϵ_{i1} that preventing one to buy.
- But in CPG, price fluctuation is fairly common (but not always, e.g., Walmart uses EDLP).
- Price fluctuations generates more incentive to stockpile – saving motive to take advantage of low price.

Application to Scanner Data

- Use IRI scanner data on laundry detergents to estimate the discount factor as well as other structural parameters.
- Use data from 2001-2007.
  - The first four years are used to construct initial inventory.
  - The final three years of data for estimation.
  - At least made 5 purchases during 2005-2007.
  - 312 households.
- Complications:
  - Many brands.
  - Even with 3 or 4 continuous state vars, the model becomes computationally burdensome if we use nested fixed point algorithm.
  - We want to incorporate persistent unobserved consumer heterogeneity in brand preferences, price sensitivities, etc.

Execution Details

- Previous research model consumption rate to be random (e.g., EIK, HN). They need to deal with unobserved inventory.
- We calibrate a household level *consumption rate* by computing the sum of total quantity over the window for which a household is observed, and dividing by total observed weeks.
- Two major advantages of this approach:
  - Gives household specific consumption rate, a major source of heterogeneity.
  - Can now impute the inventory variable.
- With 5 package sizes, allowing a flexible storage cost function significantly increases computational burden.
- We would have to track the composition of inventory and make a strong assumption on the order in which different bottles are used.
- We assume storage costs are zero until a cutoff (imagine you set aside space in your laundry room for detergent).
  - We relax this assumption and model storage costs as a function of bottles. The storage costs for that version are very close to zero.

Estimation

- Since it is infeasible to estimate a model with brand specific inventory, where consumers track brand specific prices, we apply the simplifications of Hendel and Nevo (2006).
  - All utility from brand consumption occurs at the time of purchase.
  - Inclusive value sufficiency.

- We use the Bayesian estimation method proposed by Imai, Jain and Ching (2009) to estimate the model.
  - Alleviate computational burden while incorporating persistent unobserved heterogeneity.

Highlights of IJC Method

- In the conventional approach, the value functions need to be solved at every trial parameter vector (θ^{*r}).
- The value functions computed at past parameter vectors are simply thrown away!
- Imai, Jain and Ching (2009) (IJC) algorithm:
  - In each MCMC iteration, the value function is only partially solved (at the minimum, only apply the Bellman operator once). We call them **pseudo-value functions**.
  - Store those pseudo-value functions evaluated at past parameter vectors, and use them to approximate the value functions at the current parameter vector nonparametrically.
  - This nonparametric approximation can be computationally much cheaper than the method of successive approximation.

IJC Algorithm

- Outer loop (MCMC algorithm)
  - Similar to the conventional Bayesian approach.
  - Use the likelihood constructed based on **pseudo alternative specific value functions**, $\tilde{V}_j^r$. (thus, we also call the likelihood the pseudo-likelihood).
- Inner loop (Key innovation of the IJC algorithm)
  - Approximate the expected future value at θ^{*r} by the weighted average of the past pseudo-value functions.
  - Apply the Bellman operator once to get pseudo-value function evaluated at θ^{*r}, and store it.

IJC Inner Loop

- Let $H^r = \{\theta^{*l}; \tilde{V}^l(s, p^l; \theta^{*l}), \forall s\}_{l=r-N}^{r-1}$ be the outcome of the algorithm to iteration $r-1$.

- For each s, the expected future value at the current parameter value (θ^{*r}) is approximated as

$$\hat{E}^r_{p'}[V(s, p'; \theta^{*r})] = \sum_{l=r-N}^{r-1} \tilde{V}^l(s', p^l, \theta^{*l}) \frac{K_h(\theta^{*r} - \theta^{*l})}{\sum_{k=r-N}^{r-1} K_h(\theta^{*r} - \theta^{*l})},$$

where $K_h()$ is a kernel with bandwidth $h > 0$.

- Pseudo alternative specific value functions are then

$$\tilde{V}^r_j(s, p_j; \theta^{*r}) = \begin{cases} \alpha_j - \gamma p_j + \beta \hat{E}^r_{p'}[V(s, p'; \theta^{*r})] & \text{if } s_j < \bar{S}_j - 1, \\ \alpha_j - \gamma p_j + G_j + \beta \hat{E}^r_{p'}[V(s, p'; \theta^{*r})] & \text{if } s_j = \bar{S}_j - 1, \end{cases}$$

$$\tilde{V}^r_0(s, p_j; \theta^{*r}) = \beta \hat{E}^r_{p'}[V(s, p'; \theta^{*r})].$$

IJC Inner Loop

- Simulate one draw of price vector, p^r, from the known price distribution. Apply the Bellman operator once and obtain the pseudo-value function:

$$\tilde{V}^r(s, p^r; \theta^{*r}) = E_\epsilon \max_j \{\bar{U}_{ijt}(s, p^r; \theta^{*r}) + \epsilon_{ijt} + \hat{E}_{p'}[V^l(s, p'; \theta^{*r})]\}.$$

  We store $\{\theta^{*r}; \tilde{V}^r(s, p^r; \theta^{*r}), \forall s\}$ and update the outcome to H^{r+1}.

Estimation Results

- All specifications include the following demographic variables at the household level: income, age, education and size.

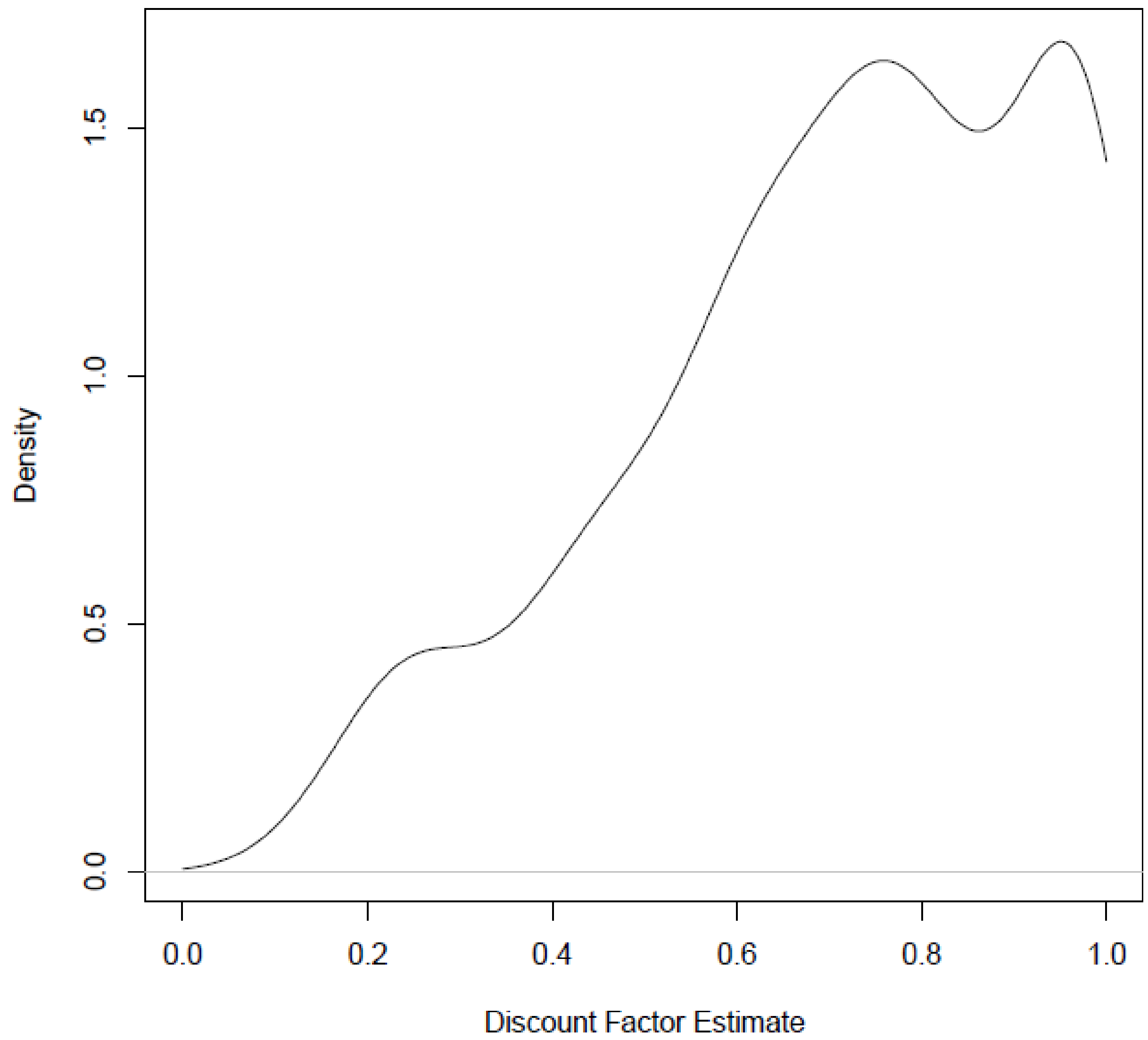
- Two specifications:
  - DIC of estimating β: 40511 (marginal log-likelihood: -19585.4)
  - DIC of fixing = 0.9995: 40773 (marginal log-likelihood: -19768.0)

Estimates

Parameter	1st Quartile	Median	Mean	3rd Quartile
Price Coefficient	-0.29	-0.24	-0.27	-0.21
	[-0.31, -0.28]	[-0.26, -0.23]	[-0.29, -0.26]	[-0.22, -0.2]
Stockout Cost	0.29	0.39	0.48	0.5
	[0.24, 0.36]	[0.31, 0.49]	[0.37, 0.66]	[0.4, 0.67]
Discount Factor	0.62	0.94	0.71	0.99
	[0.14, 0.89]	[0.81, 0.98]	[0.58, 0.82]	[0.98, 1]
Fixed Cost of Purchase	-	-	-1.83	-
			[-1.91, -1.77]	
Log-likelihood	-19585.37			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

Individual-specific Discount Factor Estimates



What have we learnt so far?

- Discount factor is heterogenous.
- Average estimated discount factor is 0.71, which is much lower than the standard calibration approach.
  - With weekly data, the standard approach (calibrating it using interest rate) is about 0.9995.
- What does 0.71 mean? It suggests that consumers plan weeks ahead when it comes to purchase decisions of laundry detergent.
- The standard approach assumes consumers plan many years ahead for laundry detergent purchase.
- You can decide which one is more reasonable.

Counterfactual

- Consider 100 oz Tide bottle.
- Promotion depth expt: Cut the promotional price by half.
- Promotional freq expt: Double the deals by randomly assign nondeal observations to be deal.
- We report changes in quantity sold and revenue in the table.

Increased Promotional Depth				
	Estimated Discount Factor		$\beta = 0.9995$	
Counterfactual	Quantity	Revenue	Quantity	Revenue
Updated Expectations	502.72	1018.48	576.95	1183.98
	[448, 560]	[777.49, 1256.67]	[514, 641]	[916.38, 1452.15]
Expectations from Data	507.56	1034.54	599.05	1255.78
	[452, 566]	[794.98, 1269.93]	[532, 669]	[972.66, 1542.56]
Increased Promotional Frequency				
	Estimated Discount Factor		$\beta = 0.9995$	
Counterfactual	Quantity	Revenue	Quantity	Revenue
Updated Expectations	49.34	232.99	55.19	276.34
	[-11, 110]	[-214.35, 679.05]	[-7, 118]	[-171.78, 734.91]
Expectations from Data	49.62	234.77	55.83	279.28
	[-11, 111]	[-213.24, 681.19]	[-7, 119]	[-172.65, 735.99]

Conclusion

- We shows the discount factor in consumer stockpiling model is identified.
  - It is highly heterogenous.
  - Its average value is much lower than the standard calibrated value.
- This can be achieved if we model storage cost more realistically.
  - Storage cost <u>can be flat </u>for an extended region of inventory.
- This leads to exclusion restrictions - a decrease in inventory does not change the storage cost and current flow utility, but it changes the expected future payoffs.
- The previous literature has imposed smoothness assumptions on storage cost, assuming it is continuous in inventory. If one allows this function to be flexible, this seemingly innocuous assumption rules out exclusion restrictions naturally arise from the stockpiling problem.
- Takeaway: Think outside the box. Do not let the existing literature to bound you.
- If you want to learn more about the estimation method by Imai, Jain and Ching (2009, Ecta), you can also refer to Ching, Ishihara, Imai and Jain (2012, QME).

Estimating Position and Social Influence Effects in Online Search

Andrew T. Ching

Johns Hopkins Carey Business School

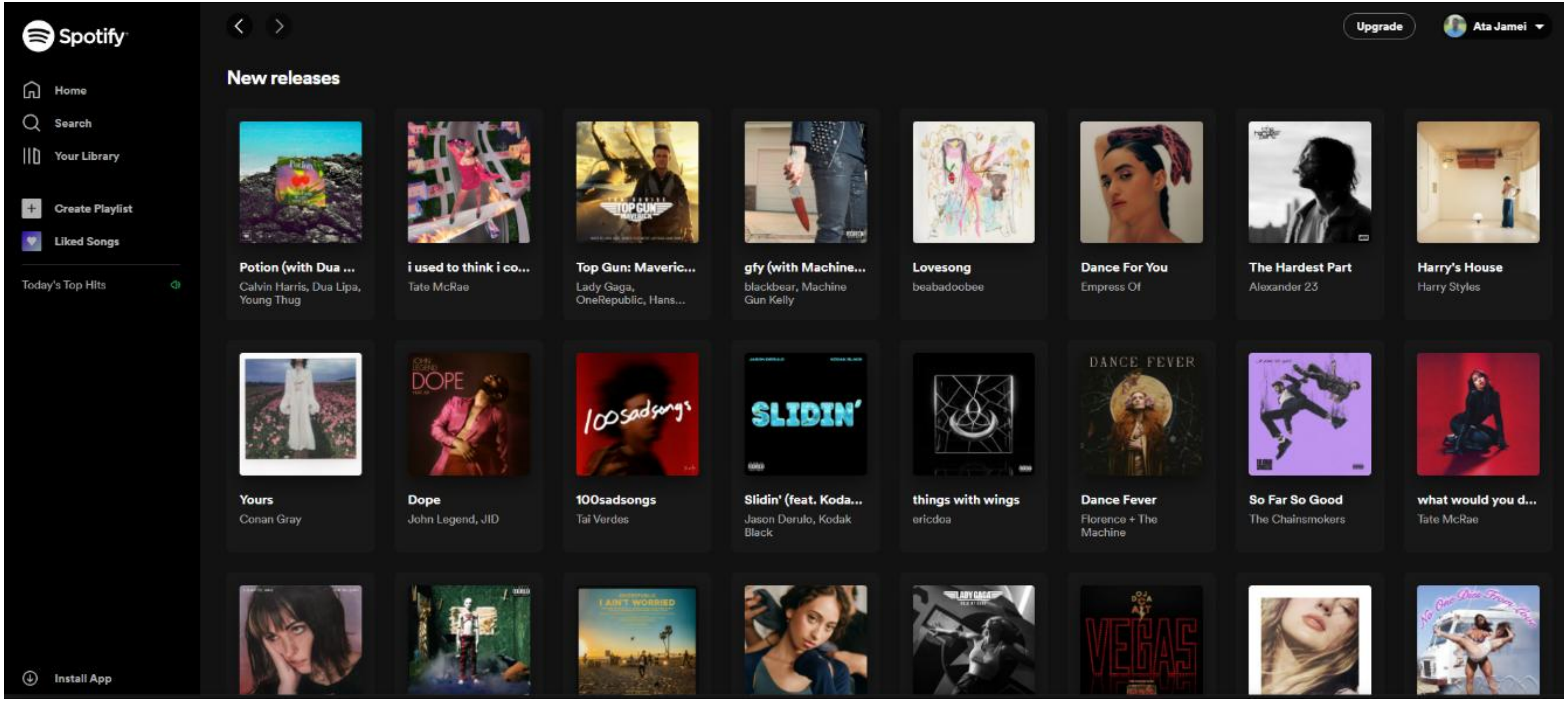
Joint work with

Ata Jameei-Osgouei[1], Shervin S. Tehrani[2], and Brian Ratchford[3]

Naveen Jindal School of Management, UTD[1,2,3]

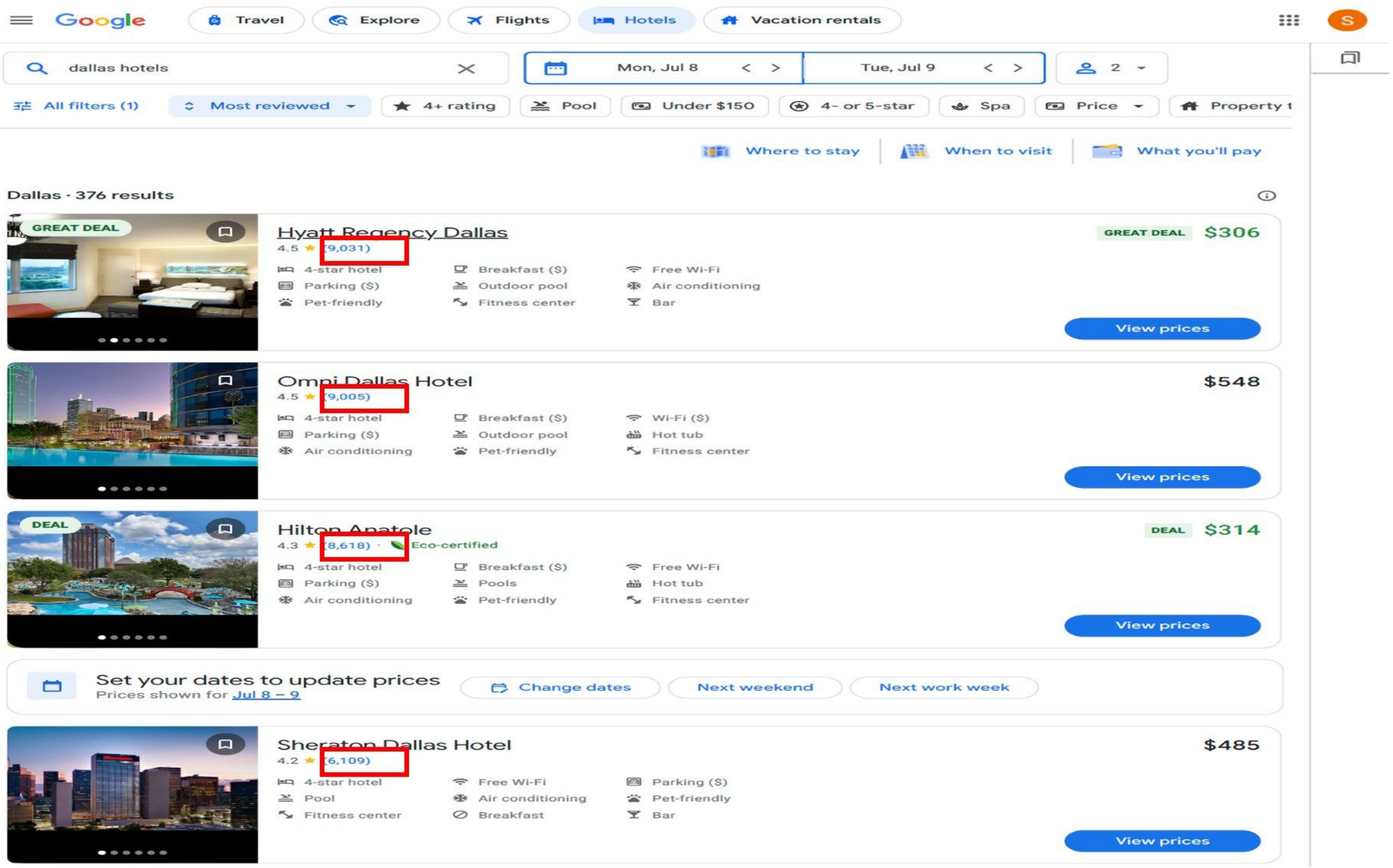
Spotify

- Spotify use a matrix display format to list musics
- It does not show any social influence, i.e., the song's popularity
- The only clue to search is the song names and artist's names



Google Hotel Search

Sorting Based Number of Reviews



Research Questions

- How should a platform display products to consumers?
  - ✓Should we disclose the social influence information (i.e., popularity)?
  - ✓Conditional on disclosing, should we sort products based on their popularity?
  - ✓Should we display products in a Column or a Matrix format?

- Consumers may change their behavior on the platform:
  - ***Search pattern, Search time, Purchase behavior***
    - ***Can change consumer welfare (or the customer value generated by platform).***
    - ***Can also affect the platform's revenue in the short-run.***

Research Questions (cont'd)

- To answer these questions, it is important to disentangle the product position and social influence (popularity) effects in consumer search behavior.

- But this is challenging because platforms usually sort products based on their popularity.

- We address this identification problem by using field experiments and an empirical structural search model.

Identification Challenge

- How do we disentangle the position and social influence effects?
- We will make use of field experiments.
  - Independent Group: Songs are randomly sorted without the popularity information.
    - This allows us to estimate the position effects.
  - Social Groups: Songs are sorted by the popularity information.
    - Fixing the position effects estimated from the Indep. group, the remaining search behavior not explained by the position effect identifies the social influence effect.

Field Experiment Data

- Music Lab created by Salganik, a large field experiment with around 10,000 participants.
- One needs to sample from 48 ***unknown songs***, and then decide whether to download them.
- We make use of three experiments conducted on this platform by Salganik and his coauthors (2006).
- Experiment 1: Matrix Display Format.
- Experiments 2 and 3: Column Display Format.
- In each experiment, there is one Independent group and several Social Influence groups. Participants are randomly assigned to them.

Field Experiment Data

Each experiment has:

- Independent group: popularity info not shown.
- Social Influence groups: popularity info is shown next to the song.

- In Experiment 1 (matrix): songs are sorted randomly in both Independent and Social Influence groups.
- In Experiments 2 and 3 (column): songs are sorted randomly in Independent group, but sorted by popularity in Social Influence group.

- Experiments 2 and 3 mimic the real world settings.

Experiment Design (Two Set of Experiments: Matrix vs. Column)

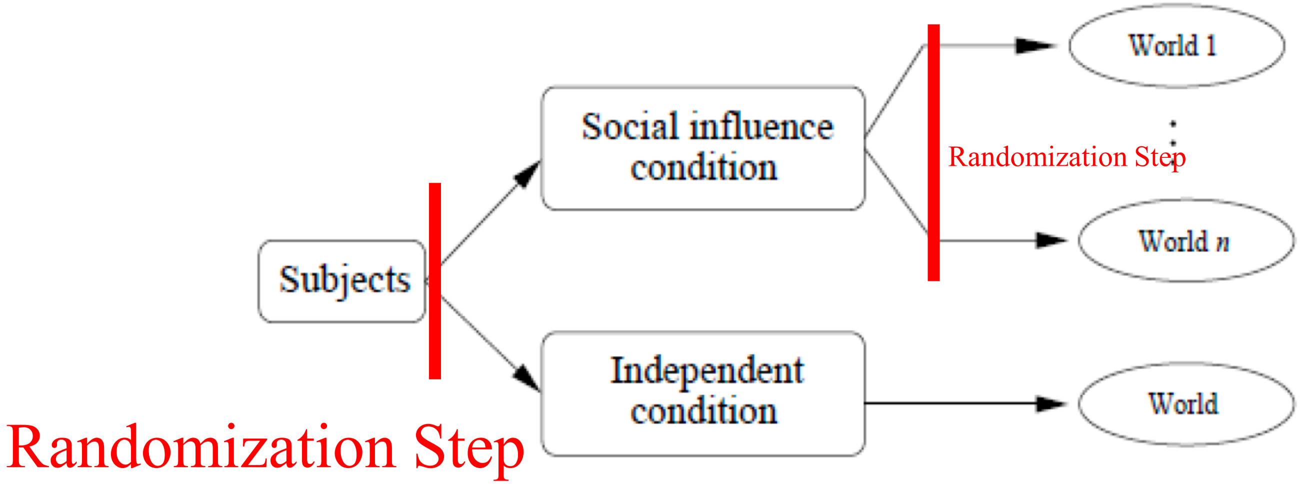
Social Group

- Each subject can observe the download numbers for each song by all previous subjects in her world.

Independent Group

- Each subject only observes songs' and bands' name (no #downloads)

Figure 1: Experiment design



Experiment Design

- ✓Each subject can sample as many songs as they want.
- ✓One can sample the entire song or just a fraction.
- ✓One can download any songs but <u>must sample it first</u>.

Experiment Design

Column Format

Social Group: Sorted based on popularity (# of downloads).

Indep. Group: Sorted randomly.

Social Group

[Help] [Log off]	# of downloads
PARKER THEORY: "she said"	85
STRANGER: "one drop"	50
FORTHFADING: "fear"	50
NOT FOR SCHOLARS: "as seasons change"	42
SHIPWRECK UNION: "out of the woods"	40
STUNT MONKEY: "inside out"	38
SIMPLY WAITING: "went with the count"	32
SILENT FILM: "all i have to say"	31
UNKNOWN CITIZENS: "falling over"	30
STAR CLIMBER: "tell me"	27
DANTE: "lifes mystery"	27
GO MOREDCAI: "it does what its told"	23
52METRO: "lockdown"	23
BENEFIT OF A DOUBT: "run away"	17
BY NOVEMBER: "if i could take you"	16
RYAN ESSMAKER: "detour_(be still)"	15
UNDO: "while the world passes"	13

Social Signal

Independent Group

[Help] [Log off]
SIMPLY WAITING: "went with the count"
SILENT FILM: "all i have to say"
UNKNOWN CITIZENS: "falling over"
STAR CLIMBER: "tell me"
DANTE: "lifes mystery"
GO MOREDCAI: "it does what its told"
52METRO: "lockdown"
BENEFIT OF A DOUBT: "run away"
BY NOVEMBER: "if i could take you"
RYAN ESSMAKER: "detour_(be still)"
PARKER THEORY: "she said"
STRANGER: "one drop"
FORTHFADING: "fear"
NOT FOR SCHOLARS: "as seasons change"
SHIPWRECK UNION: "out of the woods"
STUNT MONKEY: "inside out"
UNDO: "while the world passes"

Experiment Design

Matrix format (16*3)

Social Group: Randomly sorted

Social Group

	# of downloads	[Help] [Log off]	# of downloads		# of downloads
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HARTSFIELD: "enough is enough"	20	GO MOREDCAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink agression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
SALUTE THE DAWN: "i am error"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22

Experiment Design

Matrix format (16*3)

Indep. Group: Randomly sorted

Independent Group

	# of downloads	[Help] [Log off]	# of downloads		# of downloads
HARTSFIELD: "enough is enough"		GO MOREDCAI: "it does what its told"		UNDO: "while the world passes"	
DEEP ENOUGH TO DIE: "for the sky"		PARKER THEORY: "she said"		UP FOR NOTHING: "in sight of"	
THE THRIFT SYNDICATE: "2003 a tragedy"		MISS OCTOBER: "pink agression"		SILVERFOX: "gnaw"	
THE BROKEN PROMISE: "the end in friend"		POST BREAK TRAGEDY: "florence"		STRANGER: "one drop"	
THIS NEW DAWN: "the belief above the answer"		FORTHFADING: "fear"		FAR FROM KNOWN: "route 9"	
NOONER AT NINE: "walk away"		THE CALEFACTION: "trapped in an orange peel"		STUNT MONKEY: "inside out"	
MORAL HAZARD: "waste of my life"		52METRO: "lockdown"		DANTE: "lifes mystery"	
NOT FOR SCHOLARS: "as seasons change"		SIMPLY WAITING: "went with the count"		FADING THROUGH: "wish me luck"	
SECRETARY: "keep your eyes on the ballistics"		STAR CLIMBER: "tell me"		UNKNOWN CITIZENS: "falling over"	
ART OF KANLY: "seductive intro, melodic breakdown"		THE FASTLANE: "til death do us part (i dont)"		BY NOVEMBER: "if i could take you"	
HYDRAULIC SANDWICH: "separation anxiety"		A BLINDING SILENCE: "miseries and miracles"		DRAWN IN THE SKY: "tap the ride"	
EMBER SKY: "this upcoming winter"		SUM RANA: "the bolshevik boogie"		SELSIUS: "stars of the city"	
SALUTE THE DAWN: "i am error"		CAPE RENEWAL: "baseball warlock v1"		SIBRIAN: "eye patch"	
RYAN ESSMAKER: "detour_(be still)"		UP FALLS DOWN: "a brighter burning star"		EVAN GOLD: "robert downey jr"	
BEERBONG: "father to son"		SUMMERSWASTED: "a plan behind destruction"		BENEFIT OF A DOUBT: "run away"	
HALL OF FAME: "best mistakes"		SILENT FILM: "all i have to say"		SHIPWRECK UNION: "out of the woods"	

Remarks about the experimental data

- Like most experimental studies, Salganik et al. (2006) compares the treatment (social influence) group and control (independent) group and use the difference to provide evidence that social influence matters in determining which products succeed in the market.
  - They used Gini Index of #downloads as a dependent variable.

- But Salganik et al. (2006) made no attempt to understand the role of product position and social influence on consumer search behavior.

- We argue that their data contain much more information to allow us to address the identification challenge and disentangle these two effects.

Summary Stat for Search Sessions

Matrix Design

independent group (N=773)

	Mean	Std. Dev.	Min	Max
songs listened	6.39	8.74	1	48
songs downloaded	2.02	4.62	0	47
participant age	22.05	9.35	9	77
participant male	0.63	0.48	0	1
Internet connection (dial-up)	0.21	0.41	0	1

Social group (N=2960)

	Mean	Std. Dev.	Min	Max
songs listened	6.79	9.20	1	48
songs downloaded	2.23	4.51	0	48
participant age	22.46	9.84	9	80
participant male	0.65	0.48	0	1
Internet connection (dial-up)	0.23	0.42	0	1

Column Design

independent group (N=1803)

	Mean	Std. Dev.	Min	Max
songs listened	9.18	11.44	1	48
songs downloaded	2.15	5.37	0	48
participant age	26.10	11.35	10	82
participant male	0.47	0.50	0	1
Internet connection (dial-up)	0.16	0.36	0	1

Social group (N=4004)

	Mean	Std. Dev.	Min	Max
songs listened	7.10	9.20	1	48
songs downloaded	2.53	5.54	0	48
participant age	22.28	10.07	11	81
participant male	0.35	0.48	0	1
Internet connection (dial-up)	0.24	0.43	0	1

Preliminary Evidence

➢How social information affects sampling and downloading behavior

Figure 4: The effect of position on click-through rate and conversion rate: Column Design in Social Group

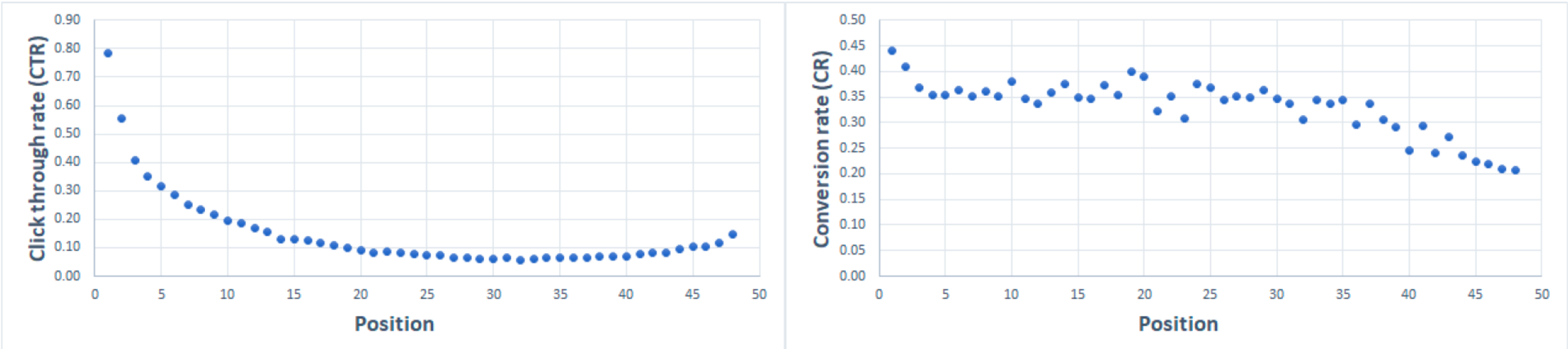
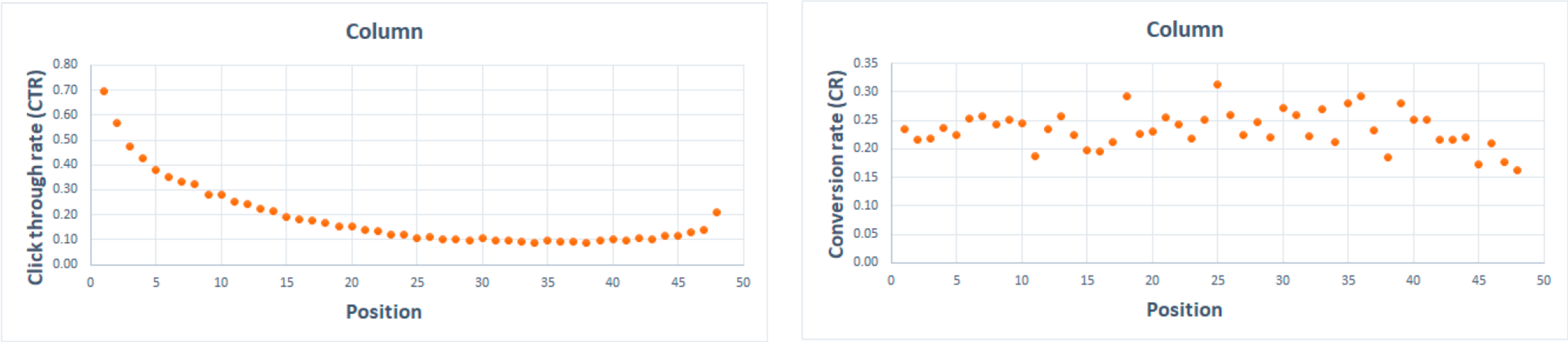


Figure 3: The effect of position on click-through rate and conversion rate: Independent Group



Model

Generalized Weitzman Model: Olszewski and Weber (2015)

- Utility depends on all the realized discovered prizes, $x_S^o = \left(x_j^o\right)_{j\in S}$: $R(x_S^o) = \boldsymbol{u(x_S^o)} - \sum_{j\in S} c_j$.
- Assumptions [drop the individual i subscript for now]:
  - Either LIKE or DISLIKE: $P(x_j^o = LIKE) = p_j, P(x_j^o = DISLIKE) = 1 - p_j$.
  - c_j is the cost of opening box j.
  - u is a concave increasing function of the total number of $LIKE$ boxes.
  - Payoff function: $R(x_S^o) = \sqrt{\sum_{j\in S} x_j^o} - \sum_{j\in S} c_j = \sqrt{k} - \sum_{j\in S} c_j$, where k *is the no. of LIKE boxes* opened so far.

- The generalized Pandora rule is optimal.
  - Selection rule: open the boxes in the order 1, 2, …, where $\frac{\boldsymbol{p_1}}{\boldsymbol{c_1}} \geq \cdots \geq \frac{\boldsymbol{p_n}}{\boldsymbol{c_n}}$, until box j.
  - Stopping rule: After finding k high quality songs, stop at box j:

$$u(k) \geq -c_j + p_j u(k+1) + \left(1-p_j\right)u(k) \quad \rightarrow \frac{c_j}{p_j} \geq u(k+1) - u(k)$$

An increasing and concave function u captures satiation.

Model

- 48 songs (in multi-armed bandit problem terminology, each song is a box).

- In ex-ante, a song's match value with a participant is either *LIKE* or *DISLIKE,* with $P(x_{ij}^{o} = LIKE) = p_{ij}$, and $P(x_{ij}^{o} = DISLIKE) = 1 - p_{ij}$.

- Participant i downloads song j iff $x_{ij}^{o} = LIKE$.

Model (cont'd)

- Search cost: $c_{ij} = \exp\left(\alpha_i + \beta\ .Position_{ij}\right)$
- $\alpha_i = \alpha_{i0} + \alpha_1 . age_i + \alpha_2 . gender + \alpha_3 . int\_connection,\ \ \alpha_{i0} \sim N(\alpha_0, \sigma_{\alpha 0}^2)$
- Utility function: $\psi_{it} = \sqrt{(\sum_{\tau=1}^{t} x_{i\tau})}$ & ψ_{it} is **increasing and concave**.
- Perceived Probability of LIKE: $p_{ijt} = \dfrac{\exp(\gamma_0 + \gamma_1 \boxed{feedback_{it}} + \gamma_2 \boxed{appeal_j} + \gamma_3 \boxed{popularity_{ij}})}{1 + \exp(\gamma_0 + \gamma_1 feedback_{it} + \gamma_2 appeal_j + \gamma_3 popularity_{ij})}$
- Download iff LIKE

Number of downloads for song j that participant i observes

Song appeal effect: song title, band name

participant i's experience from sampling songs on the experiment on her belief

$$feedback_{it} = \frac{TP_{it} - TN_{it}}{TP_{it} + TN_{it}}$$

$$TP = \#songs\ downloaded, TN = \#songs\ not\ downloaded$$

$$Appeal_j = l_j / \sum_k l_k$$

l_j: number of times song j is sampled in the independent group

Remarks on the Model

- Identification assumptions:
  - Popularity and feedback only affect the perceived probability of the match.
  - Position only affects the search cost.

- Even if we include position in the perceived probability of the match, it is still possible to apply 2-step procedure.

- $popularity_{ij}$ remains unchanged during a participant i's search session, but it changes across i.
- $feedback_{it} = \frac{TP_{it} - TN_{it}}{TP_{it} + TN_{it}}$, changes at each search step t, and captures adaptive learning (e.g., Doraszeleski, Lewis and Pakes, 2018; Li and Ching, 2024).

Estimation Strategy and Identification

Social Group Data

- Matrix Format Experiment – One-step Estimation Approach
  - Because songs are always randomly sorted in this experiment, we can apply standard one-step estimation to estimate all parameters at the same time for each condition (Indep. vs. Social)
- Column Format Experiments (resembles real world settings) – Two-step Estimation Approach
  - Step 1: Estimate all parameters using the Independent group data. Without social influence information, the position effect is identified.
  - Step 2: Fixed the position effect and its related parameters (parameters for search costs) at Step 1, estimate the rest of the parameters using the Social Influence groups data.
    - The search behavior not explained by the position effect (or search costs) identify the social influence effect and other pre-search belief parameters.

Estimation Results

Table 3: Estimation Results

	Matrix			Column		
	Independent	Social Influence		Independent	Social Influence	
		Two-Step	One-Step		Two-Step	One-Step
Search Cost						
α_0	-1.54^{***}	-1.54^{***}	-1.52^{***}	-1.59^{***}	-1.59^{***}	-1.48^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
σ_{α_0}	0.37^{***}	0.37^{***}	0.41^{***}	0.41^{***}	0.41^{***}	0.43^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\alpha_1(age)$	-0.002^{**}	-0.002^{**}	-0.002^{***}	-0.003^{***}	-0.003^{***}	-0.002^{***}
	(0.0010)	(0.0010)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
$\alpha_2(gender)$	-0.06^{***}	-0.07^{***}	-0.06^{***}	-0.04^{***}	-0.04^{***}	-0.03^{***}
	(0.02)	(0.02)	(0.012)	(0.01)	(0.01)	(0.010)
$\alpha_3(internet)$	0.27^{***}	0.27^{***}	0.19^{***}	0.22^{***}	0.22^{***}	0.22^{***}
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
$\beta(position)$	0.0006^{***}	0.0006^{***}	0.0005^{***}	0.0010^{***}	0.0010^{***}	0.0001^{*}
	(9.2E-05)	(9.1E-05)	(5.3E-05)	(5.2E-05)	(5.2E-05)	(6.2E-05)
Pre-search Belief						
γ_0	0	0	0	0	0	0
$\gamma_1(feedback)$	1.24^{***}	1.20^{***}	1.17^{***}	1.28^{***}	1.19^{***}	1.05^{***}
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.01)
$\gamma_2(appeal)$	0.008^{***}	0.007^{***}	0.006^{***}	0.005^{***}	0.003^{***}	0.004^{***}
	(0.002)	(0.000)	(0.001)	(0.002)	(0.000)	(0.001)
$\gamma_3(popularity)$		0.009^{***}	0.009^{***}		0.019^{***}	0.029^{***}
		(0.000)	(0.001)		(0.001)	(0.001)
Log-Likelihood	-5098	-19695	-19683	-12637	-25602	-25284
Observations	37104	142080	142080	86544	192192	192192

Note: Robust standard errors are shown in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Estimation Results: Main Takeaway

- Column display format (position and popularity are confounding in the social influence group)
  - 2-step estimation results: Position effect estimate is 0.001, Popularity effect estimate is 0.019. Both are +ve and significant.
  - 1-step estimation results: Position effect becomes insignificant, Popularity effect increases to 0.029.

- Matrix display format (there is no confounding/identification problem)
  - 1-step estimation results: Position and Popularity effects are both +ve and significant.
  - Although there is no need to implement 2-step estimation results, we implement it as well as a robustness check.
  - 2-step estimation results: As expected, the point estimates for all parameters are very similar to those in the 1-step estimation results.

Counterfactuals

- For each participant, we keep their arrival order the same and randomly draw 100 α_{i0}.
- Note that in the actual experiment, each participant is only assigned to one condition. In the counterfactual experiment, we can predict how each participant would behave in all possible scenarios.
- Four counterfactuals:
  - CF1: Songs are sorted randomly and no popularity information
  - CF2: Songs are sorted randomly with popularity information displayed.
  - CF3: Songs are sorted according to displayed popularity information.
  - CF4: Popularity displayed and songs are sorted individually based on each participant's predicted download likelihood.
- Note that the followings are new:
  - (i) CF2 in column format.
  - (ii) CF3 in matrix format.

Counterfactuals

- One challenge is we do not observe the ex-post download decisions of songs which were not sampled by a participant in the experiment.
  - We do not observe whether participants like the songs which they haven't sampled in the data.
- People have heterogeneous tastes for music.
  - Need to take into account several features and possible hidden layers to obtain a reliable prediction of ex-post download probabilities.
- We predict the counterfactual probability of LIKE by using *Content-based Filtering*.

Counterfactuals

- 8 Scenarios (2 popularity display format per CF)
  - CF1 : Random assortment, no popularity
  - CF2 : Random assortment, with popularity
  - CF3 : Sorted with popularity
  - CF4: Sorted based on personalized preference

Popularity: downloads displayed for each participant.

Download probability: Predicted by content-based filtering.

- Average #sample (#listens), #downloads, search efficiency (= #downloads/#sample) per search.

	Matrix			Column		
	Samples (l)	Downloads (k)	Efficiency (w)	Samples (l)	Downloads (k)	Efficiency (w)
CF1: **Randomly Sorted without Popularity**	7.76 (0.112)	1.29 (0.02)	0.303 (0.003)	8.65 (0.116)	1.50 (0.02)	0.293 (0.002)
CF2: **Randomly Sorted with Popularity**	8.26 (0.108)	1.38 (0.02)	0.332 (0.004)	9.36 (0.109)	1.63 (0.02)	0.336 (0.004)
CF3: **Sorted based on Popularity**	7.90 (0.098)	1.44 (0.02)	0.348 (0.006)	8.98 (0.109)	1.72 (0.03)	0.357 (0.006)
CF4: **Sorted based on Personalized Preference**	7.13 (0.11)	1.76 (0.02)	0.436 (0.003)	8.14 (0.11)	2.18 (0.03)	0.454 (0.003)

Note: Please refer to Web Appendix D for an explanation of how l, k, w are calculated. Robust standard errors are shown in parentheses. The t-test reveals significant differences between the relevant variables.

Conclusion

- Take advantage of an experimental design to disentangle product position and social influence effects in consumer search behavior.
- The column list design works better conditional on revealing the popularity:
  - ✓ ***Disclosing social information using random assortment*** *leads the* ***highest sampling rate***.
  - ✓ ***Disclosing social information using sorted social information*** *leads the* ***highest downloading rate and search efficiency***.
- Experimental data can allow us to learn much more than simple DID design, when combining it with a structural model.
  - Other examples include Chan and Hamilton (2006), Hamilton, McManus, Pantano, Trogdon (2025), Misra and Nair (2011), etc.

Thank You!