What are news readers willing to pay for?

Using Stan for structural econometrics

Greg Martin, Cameron Pfiffer, Shosh Vasserman

Stanford GSB

June 23rd, 2023

What is the future of news?

What is the future of news? (as of 2016)

PBS NEWS HOUR

Left: Buzzfeed employees work at the company's headquarters in New York. BuzzFeed has come a long way from cat lists. The kittens haven't disappeared, but these days there is serious journalism as well. New technology that has upended how news and advertising are produced and distributed. Photo by Brendan McDermid/Reuters

By – Jeffrey Dvorkin

Column: Why click-bait will be the death of journalism

Economy Apr 27, 2016 6:11 PM EDT

Journalism is (once again) in crisis. This time, the sky really does seem to be falling.

Leave your feedback

Share

Go Deeper

buzzfeed

canada

freelance

gig economy

internet news

making sen$e

making sen$e columnist

media

uber

What is the future of news? (as of 2023)

TheDesk.net
10TH ANNIVERSARY

STREAMING TELEVISION RA

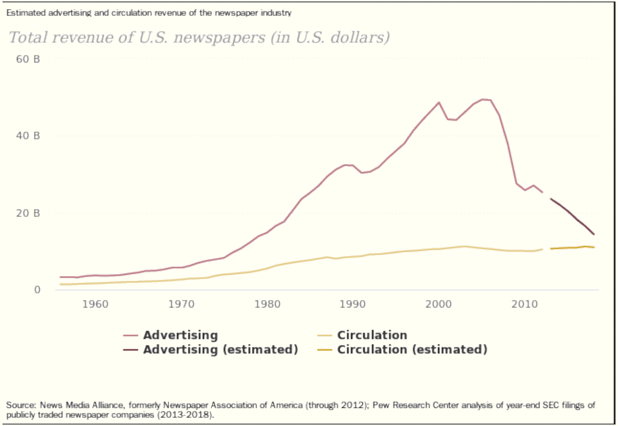
BuzzFeed says clickbait revenue not enough to support news business

BuzzFeed's CEO admitted BuzzFeed News wasn't profitable, and apparently would never be.

By: Matthew Keys | April 20, 2023

BuzzFeed is shutting down its news division and laying off around 180 workers after executives at the

Ad revenue is declining



Our question

- What should local news groups do to become more profitable?
  - What types of articles drive *clicks* (ad revenue)?
  - What types of articles drive *subscriptions* (circulation revenue)?
- Understanding this requires understanding reader preferences
- Equipping news agencies with a thorough understanding of reader preferences will:
  - Help them adjust their content mix to optimize for subscription revenue
  - Help them adjust their offers (price/duration) to optimize for subscription revenue
  - Help communities maintain access to local news
- This is a work in progress! Model development is still underway.

Out Stan use

- We use Stan for MCMC sampling
- BridgeStan and Julia for mode estimation (due to Jacobian adjustment)
- Stan's (reasonably) efficient parallelism made cluster computing fast, just throw MORE CORES at the problem
- Stan is a pretty good tool for arbitrary statistical models, not just Bayes stuff

Our data

- Universe of online activity from a mid-sized local newspaper
- Full coverage for January 2020-March 2023
    - Today we are restricting to September 2020-October 2020
- Tens of thousands of articles with full text & bylines.
- Complete Google Analytics visit data at user $\times$ day $\times$ article level.
    - Paywall events with menu of subscription offers presented
    - What the user is trying to read when paywalled
    - Location, adblocker status, fingerprint, etc.
- Subscription history of individual users.
    - Offer price and terms accepted, if any
    - Reading history before and after subscription
    - Renewal decision (inferred)
- Article features: author identity, localness, topic, investigative score (Turkel et al, 2021)
- Soon: OpenAI embeddings of article text

The site

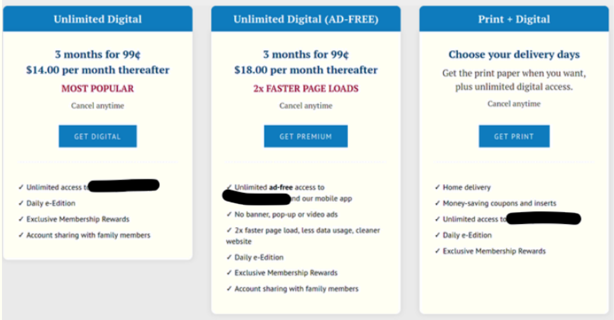
- Users arrive to the site
- We observe all user actions
- We observe the actions of the site, e.g. paywalls, offers presented, etc.
- We want to understand user's preferences + willingness-to-pay for different types of content

How paywalls work

- Readers get a certain number of free articles per month, but do not know how many
- Free article counts can change over time and are not announced
- Once paywalled, a reader cannot read unless they
  - (a) subscribe or
  - (b) wait until the next month
- Paywalls are stochastic from a user's perspective
- Quasi-randomness + prompt for a purchasing decision $\approx$ an identification strategy for preferences!

Menus

If the reader hits their free article limit, they are paywalled and presented with a menu of subscription offers:



Offers

- Offers have an introductory price and length (e.g. \$0.99 for 3 months)
- After the introductory period, the price increases (e.g. \$14 per month)
- Some offers are ad-free, in that they disable ads on the website
- We define an individual price/length combination as an *offer*.
- A combination of one or more offers is a *menu*.
- Lots of unique deals on national holidays

Our plan

1. Derive features of each article that describe their content (section, local-focus, etc.)
2. Fit a model of arrivals to each article as a function of article features
3. Use quasi-experimental variation in paywall events to estimate subscription value

Paywall propensity

- Encountering a paywall is correlated with a user's interest in the site.
- We construct a logistic paywall propensity score to receive at least one paywall during their visit based on user readership history as of each visit

$$\rho_{it} = \Lambda(\phi' z_i t)$$

- where z_{it} is a vector of user i's
  - paywall history up to day t
  - average article reading depth
  - total number of words read in the previous five weeks

Paywall propensity

- Roughly 85% of user-days have a negligible paywall propensity
- Readers are sorted into a low propensity bin ($\rho_{it} \in [0.05, 0.16)$) and a high propensity bin ($\rho_{it} \in [0.16, 1]$)
- Readers with negligible propensity are discarded (for now)

Key data features

- Readers come for lots of different types of content
- Very few subscribe
- Price is important
- Popular articles don't necessarily induce higher subscriptions

The model

The modelling goal

1. Articles of different types are produced stochastically.
2. The newspaper is more popular on some days than others.
3. Readers of different paywall propensity types

- Read different articles
- Are paywalled at different probabilities
- Get a random menu of offers

4. Readers choose whether to subscribe based on

- Base subscription value (the value of all future reading)
- Offer price + duration + other offer features
- Subscription popularity on that day, correlated with general popularity

Model overview

- Unit of observation is a single user-session interaction
  - Users are $x_i \in \{1, 2\}$, where 1 indicates a low paywall propensity type, 2 indicates a high propensity type
  - Article features are denoted X_a
- Jointly model **arrival** and **subscription** as separate (but correlated) decision processes
- In short – it's a two-step discrete choice model with some extra bells and whistles

Arrival model

- Need to account for arrivals + subscriptions jointly, since subscription rates correlate negatively with traffic
- Users have a probability of arriving to each article

$$P_{x,d,a} = \frac{\exp(\eta_x + \eta_d + \theta_x' X_a)}{1 + \sum_{a'} \exp(\eta_x + \eta_d + \theta_x' X_{a'})}$$

- η_x and η_d are user-bin and day IID shocks
- θ_x are article feature coefficients specific to each user type x
- The outside option (not visiting the site) is normalized to one

Arrival model

- Arrivals to each article are multinomial on a given day

$$N_{x,a,d} \sim \text{Binomial}(P_{x,d,a}, \overline{N})$$

- $\overline{N}$ is the maximum possible visitors on an article-day
- Assume $\overline{N}$ is twice the largest number of unique page visits observed
- Treat $N_{x,0,d}$ as the outside option

Subscription model

- Users can be paywalled or not, with a probability that varies by user type
- Paywalls come with 1-3 offers for a subscription, assume menus are exogenous
- Menus have T_o days of access for offer o, at price p_o.
- Users consider each offer presented and choose whether to purchase offer o

Subscription likelihood

Users of type x trying to read article a expects offer o to yield

$$u_{x,o,d} = (1 - I^{\text{paywall}})\beta_x' X_a + \varsigma_x I_o^{\text{adfree}} - \alpha_x p^*_{x,o,d} + \nu_x + \kappa_d + \epsilon_{i,a,d,o}$$

- where
  - I^{paywall} indicates whether a user was paywalled
  - I_o^{adfree} indicates whether offer o is ad-free (i.e. disables ads on the site)
  - $\epsilon_{i,a,d,o}$ is a type-1 extreme value shock (Gumbel error)

Average price

- $p^*_{x,o,d}$ is the average daily offer price

$$p^*_{x,o,d} = \begin{cases} p_o & \text{if} \quad T_o \geq 360 \\ (1-w)p_o + \rho w R_o & \text{else} \end{cases}$$

- Daily introductory price p_o
- Introductory length (in days) T_o
- Daily renewal price after the introductory offer ends R_o
- Price weight is $w = (360 - T_o)/360$

Parameters

Users of type x trying to read article a expects offer o to yield

$$u_{x,o,d} = (1 - I^{\text{paywall}})\beta_x' X_a + \varsigma_x I_o^{\text{adfree}} - \alpha_x p_{x,o,d}^* + \nu_x + \kappa_d + \epsilon_{i,a,d,o}$$

- where
  - β_x is a vector of article feature preferences
  - ς_x is the ad-free value
  - α_x is price disutility
  - ν_x and κ_d are user-bin & day-level IID shocks
  - Arrival and subscription parameters are correlated

$$[\theta_x, \beta_x] \sim N(0, \Sigma_{\theta,\beta})$$
$$[\nu_d, \kappa_d] \sim N(0, \Sigma_{\eta,\kappa})$$

Subscription likelihood

- Using Type 1 extreme value for $\epsilon_{i,a,d,o}$ gives us subscription probabilities for each offer o
- Subscription probabilities follow

$$P(\text{subscribe}_{x,o,d}) = \frac{u_{x,o,d}}{1 + \sum_{o'} u_{x,o',d}}$$

- Denote actual decision with binary vector $\mathbf{I}_{x,d}^{\text{subscribed}} = [0, 1, 0, 0, \ldots, 0]$
- Likelihood is then

$$\mathbf{I}_{x,d}^{\text{subscribed}} \sim \text{Multinomial}(1, \mathbf{P}(\text{subscribe}_{x,d}))$$

Quick model recap

- Arrival

$$
\begin{aligned}
N_{x,a,d} &\sim \text{Binomial}(P_{x,d,a}, \overline{N}) \\
\eta_x &\sim N(-,-)
\end{aligned}
$$

- Arrival value coefficients

$$
\begin{aligned}
[\nu_d, \kappa_d] &\sim N(0, \Sigma_{\eta,\kappa}) \\
\Sigma_{\eta,\kappa} &= \sigma'_{\eta,\kappa} \Omega \eta, \kappa \sigma_{\eta,\kappa} \\
\Omega_{\eta,\kappa} &\sim \text{LKJ}(2) \\
\sigma_{\eta,\kappa} &\sim \text{InverseGamma}(2,3)
\end{aligned}
$$

Quick model recap

- Subscription

$$\mathbf{I}_{x,d}^{\text{subscribed}} \sim \text{Multinomial}(1, \mathbf{P}(\text{subscribe}_{x,d}))$$

$$\alpha_x \sim N(-,-)$$

$$\varsigma_x \sim N(0,1)$$

- Subscription value coefficients

$$[\theta_x, \beta_x] \sim N(0, \Sigma_{\theta,\beta})$$

$$\Sigma_{\theta,\beta} = \sigma'_{\theta,\beta}\Omega\theta, \beta\sigma_{\theta,\beta}$$

$$\Omega_{\theta,\beta} \sim \text{LKJ}(2)$$

$$\sigma_{\theta,\beta} \sim \text{InverseGamma}(2,3)$$

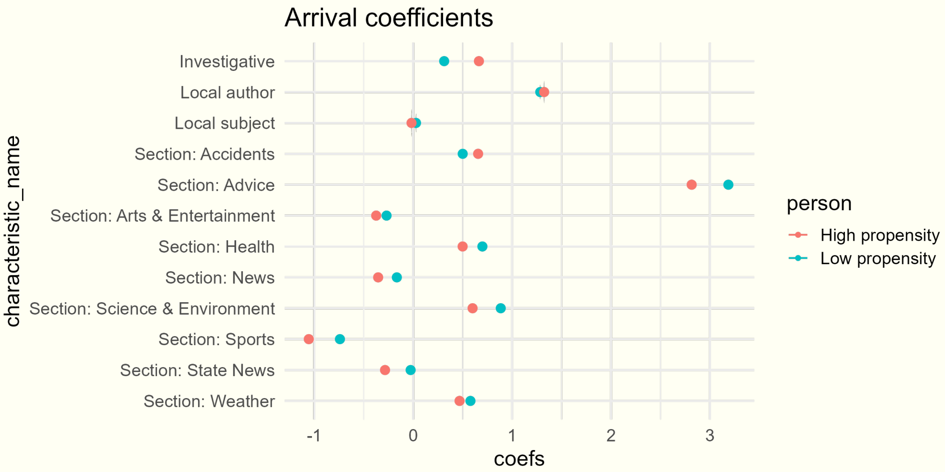
- Weighting on renewal price

$$\rho \sim \text{Uniform}(0,1)$$

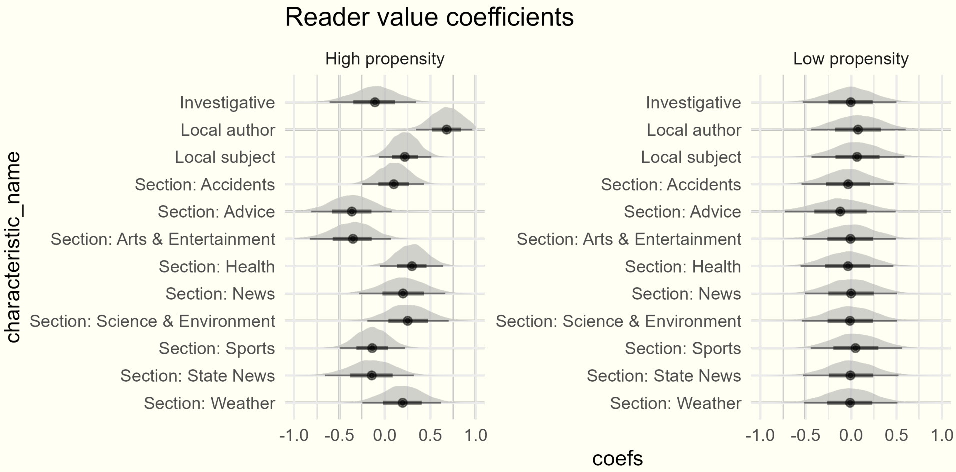
Model estimates (preliminary)

- Limit sample to September 2020 through October 2020
    - Observations in the millions (cannot report exact numbers due to NDA)
    - Very few subscriptions in the low propensity group
- 2,000 draws in Stan, 1,000 warmup draws
- Samples drawn with the Yens cluster at Stanford GSB
- 3 chains (1 failed for hardware reasons)
- Maximum rhat is 1.08
- ESS between 3,000 and 4,000 for most of the well identified parameters (i.e. not the low propensity people)

Arrival coefficients



Subscription value coefficients



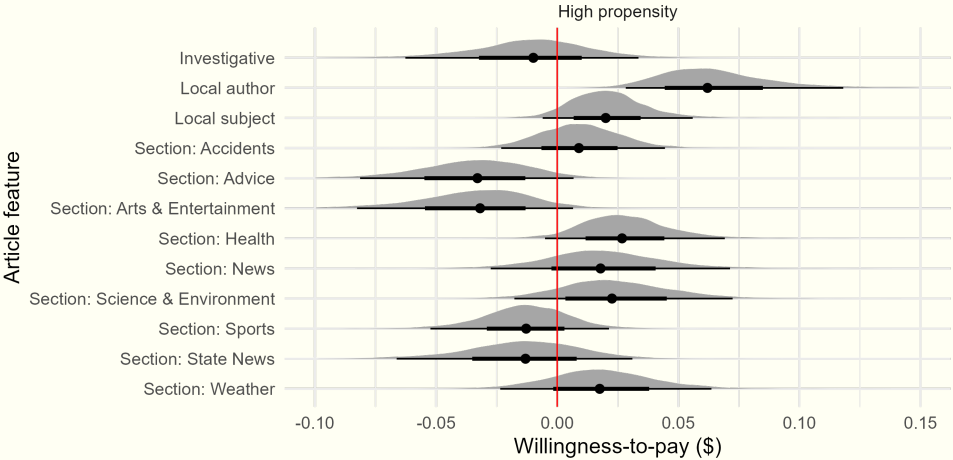
Willingness-to-pay definition

- We commonly define **willingness-to-pay** as the benefit in dollars a consumer receives from a given feature in article a ($X_{a,k}$)
- The calculation for this is commonly

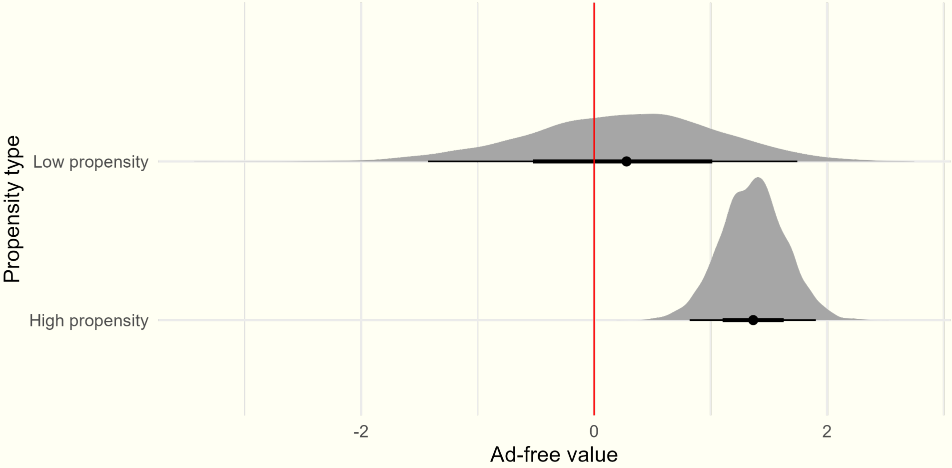
$$\text{WTP}_{x,k} = \frac{\beta_{x,k}}{\alpha_x}$$

- This follows after taking the total derivative w.r.t. price and characteristic k
- See Train's Discrete Choice book for more info
- **Intuition:** For $\beta_{x,k} = 2$ and $\alpha_x = 10$, the user would be willing to pay $\text{WTP}_{x,k} = 0.20$ dollars for an article with a 1-unit increase in article a's characteristic k

Subscription value coefficients



Does ad-free matter?



Stan usage

Stan usage

- Performance notes for bigger data sets
- Vectorization matters a lot
- `reduce_sum` is fast, `map_rect` might be faster in cluster environs (any comments on that?)
- GPU support is worthwhile if you have access to them & have reasonably large matrix operations
- Squish to simpler distributions when you can (sometimes you can't)
  - Binomial instead of Bernoulli
  - Multinomial insted of Categorical

Logit implementation notes

- Here's some thoughts on Stan implementations of logit and logit-like discrete choice models
- Logits are highly prone to `exp` overflows, especially in weird parameter spaces
- Use a good initialization
- Estimate the logit error scale parameter and rescale your utility by it, or

Comments for the Stan community

- PPLs are *excellent* statistical model prototyping tools
- Economics has lots of complex, hand-rolled models that are often sequentially improved
- Lots of these folks use maximum likelihood and the Hessian for standard errors
- Stan should consider making Hessians a bit faster + easier to access for people prototyping models

Next steps

- Estimating counterfactuals – changing the articles produced to maximize news group revenue using posterior predictive
- Write a model that better captures expected reading value
- Estimate random effects on coefficients