ECON-C5100 Digital Markets

Iivo Vehviläinen

September 7, 2022

Aalto University
iivo.vehvilainen@aalto.fi

Lecture 2: Preferences

Last time

- The square and the tower
  - Market institutions matter
- Digital marketplaces increase efficiency
  - Reduction of search costs and other frictions
  - Trade-off between privacy, discrimination, and efficiency

Today

- Preferences online
- Estimation and prediction

Recap: Market demand builds on preferences

Figure. Demand of a good depends on the preferences of the individuals who participate to the market.

But sellers have constraints

- Information: preferences of an individual are unknown
- Can only set one list price (to be discussed)

Preferences online

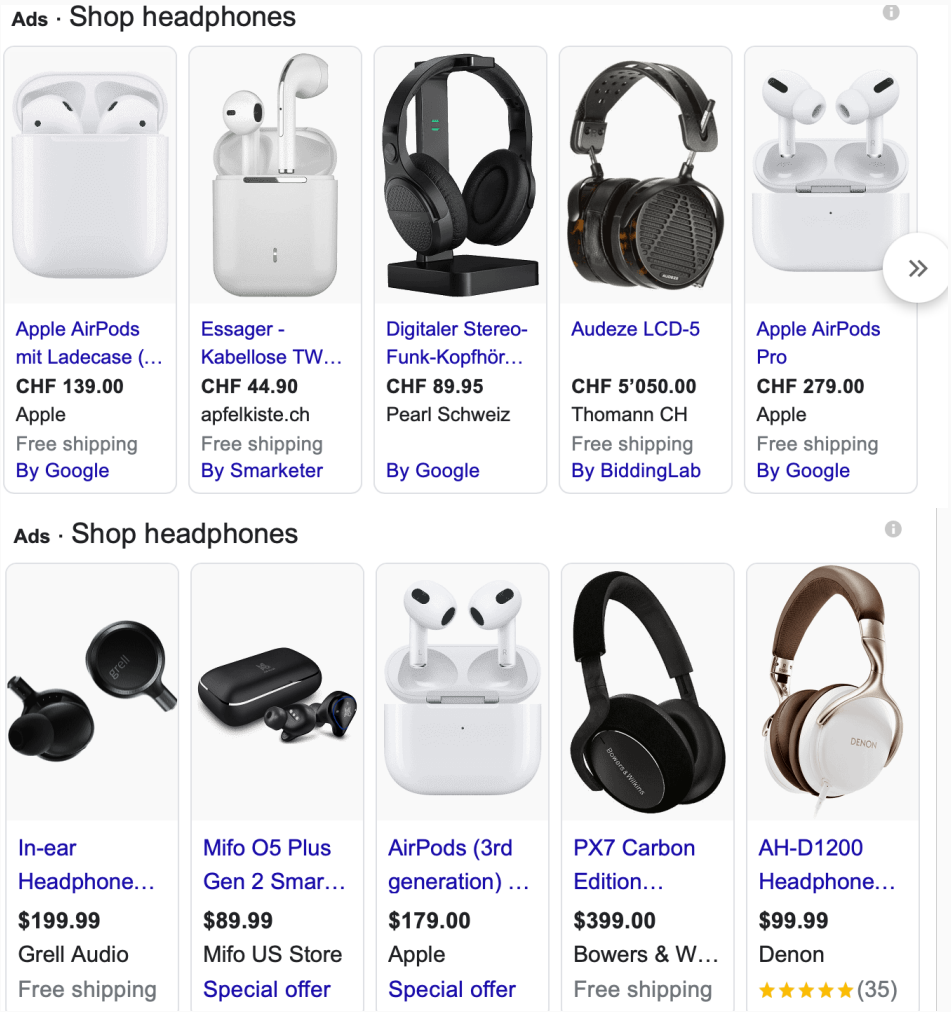


Figure. Two identical Google searches, one done from Zurich, Switzerland (top) and another from Mexico City, Mexico (bottom)

Experiment: Search engine

Google *Beats headphones*

- Report the price from the first ad that is shown to you at *Search* chat at presemo.aalto.fi/digimar
- You can use your preferred other search engine, if Google is out of the question

Use of data

- What you want to do
  - For example buy a phone, watch a movie, connect with friends
- What a firm wants to do: find out what you want to do by tracking your actions... and try to influence behavior:
  - Design a mechanism to discover your preferences
  - Collect and use data to improve performance

Example: Tiktok

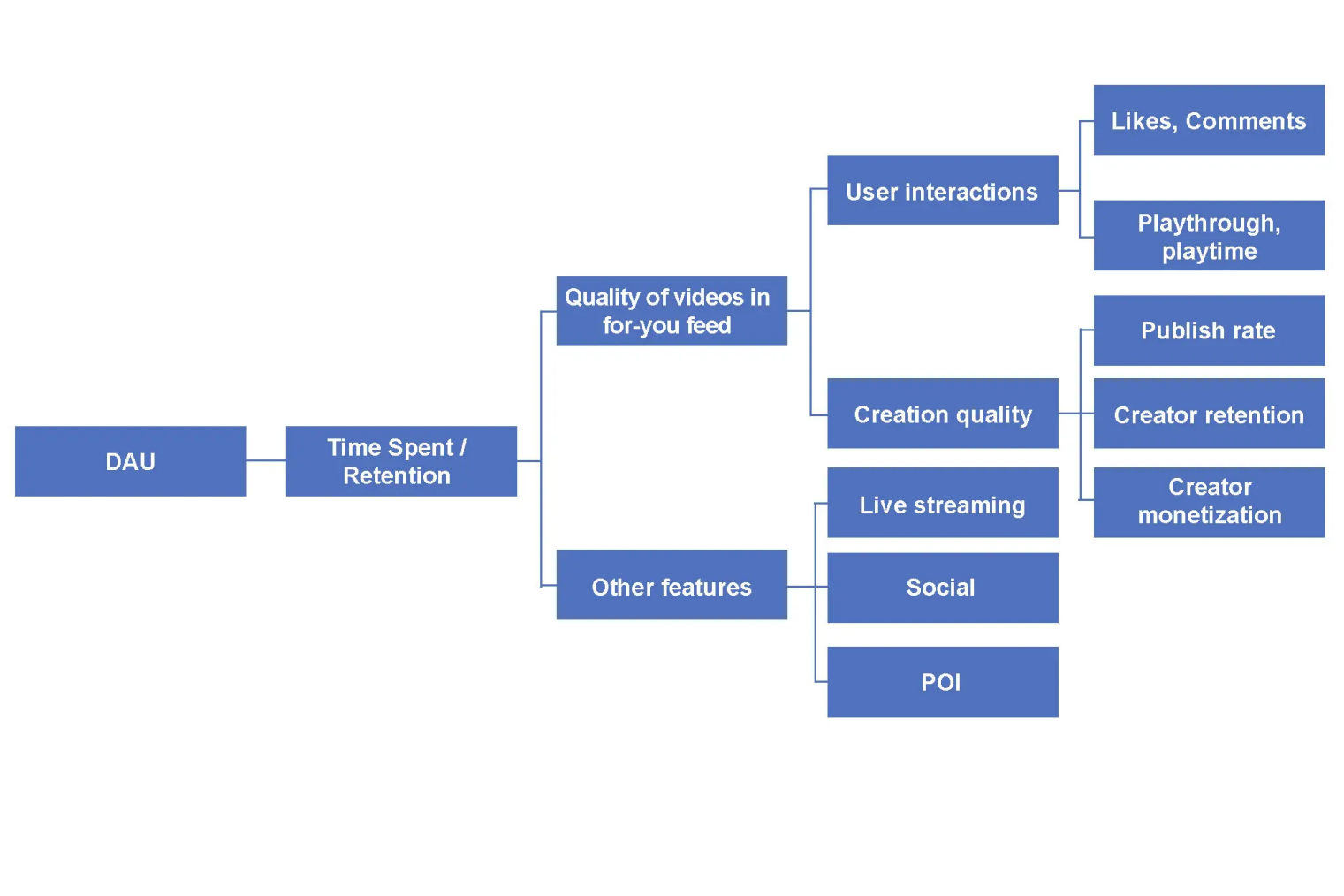


Figure. The goals of Tiktok's algorithm (DAU = daily active users)

Figure: How TikTok Reads Your Mind, New York Times, 5 Dec 2021

Tracking user data

- Big tech companies and others collect data of their users and use it to advance their business
- Ownership of this data may not be clearly defined or understood; nor is its value
- People exchange data of varying value against payment in the form of service(s) that they use

Example: Netflix

- Netflix has a 60 PB+ of data collected from all its operations, with 100+ million subscribers (around 600 MB per customer).
- Company policy is to only do data driven business decisions:
  - Recommend films
  - Choose which content to purchase
  - Create content (House of Cards)
  - Improve user experience through A/B testing
  - Give users information on which Internet provider to use
- But where should the limits on data use be?
  - Facebook was caught providing Netflix access to its user data without the consent of the users (NYT, 18 Dec 2018)

Source: Keijo Helanko / Aalto SCI, 2018.

Example: Facebook

Figure. Categories of data that Facebook collects of you.

Figure: Facebook.

Data collection

- Browser cookies are an early invention to track users
- They are still widely used, but their benefit has been reduced
  - Regulation made companies to disclose the use of cookies in the EU ("This website uses cookies to...")
  - Browsers have started to compete on privacy dimension, you can opt-out from cookies
- The tracking industry is responding with new technologies
  - Keeping people signed in to their service on a device
  - Merging user information from e.g. mobile phone locations and browsing behavior to other data
- Also, with combined data from all users, it is possible to predict your behavior, even if you don't disclose it yourself

---

Bergemann et al. 2019.

Recovering preferences from data

- There are two key motivations to find out your preferences:
  1. Understand the drivers behind your choices
  2. Predict how you will behave in a given situation
- The first part, causal inference, is a traditional field of economics (microeconometrics)
- The second part, prediction, is what the big tech firms are doing with the availability of big data

Estimation and prediction

To pinpoint the differences:

- Let's consider a simple model

$$Y_i = f(X_i) + \epsilon_i$$

- X_i are data (e.g. location, past clicks).
- Y_i are variables of interest (e.g. click to an ad).
- Function f maps values of X to Y and ϵ_i are the random noise (in measurement, unobserved variables, etc.).

Estimation and prediction

The goals in estimation and prediction are different:

1. The goal of regression estimation is unbiasedness:

$$E[\hat{f}] = f.$$

2. The goal of prediction is to minimize prediction error:

$$\hat{f} = \min_{f \in \mathcal{F}} L(f) = \min_{f \in \mathcal{F}} E[\ell(f(x), y)],$$

where e.g. $\ell(f(x), y) = (y - f(x))^2$.

Susan Athey, 2018.

The bias-variance tradeoff

If $E[\epsilon] = 0$, the expected prediction error is:

$$\begin{aligned} \text{error} &= E[(y - \hat{f}(x))^2] \\ &= E[(y - f(x))^2] + (E[\hat{f}(x)] - f(x))^2 \\ &\quad + E[(\hat{f}(x) - E[\hat{f}(x)])^2] \end{aligned}$$

for some x out of sample. Using $\epsilon = y - f(x)$, we have

$$\text{error} = Var(\epsilon) + Bias^2 + Var(\hat{f}(x))$$

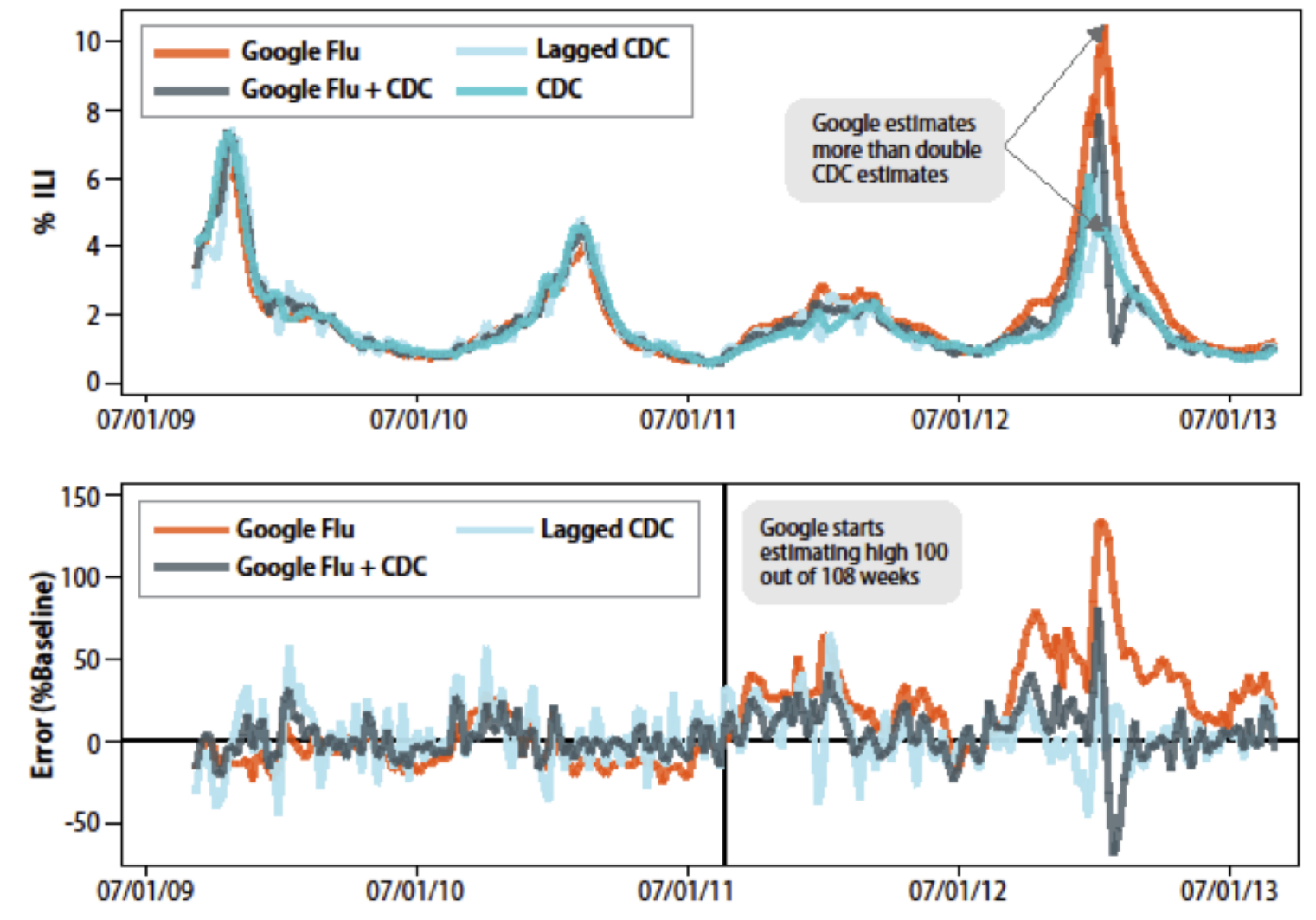
i.e. a combination of noise or variance in data ($Var(\epsilon)$), bias in estimation, and variance from the fact that a sample is used to estimate $\hat{f}$

Prediction

Basic idea:

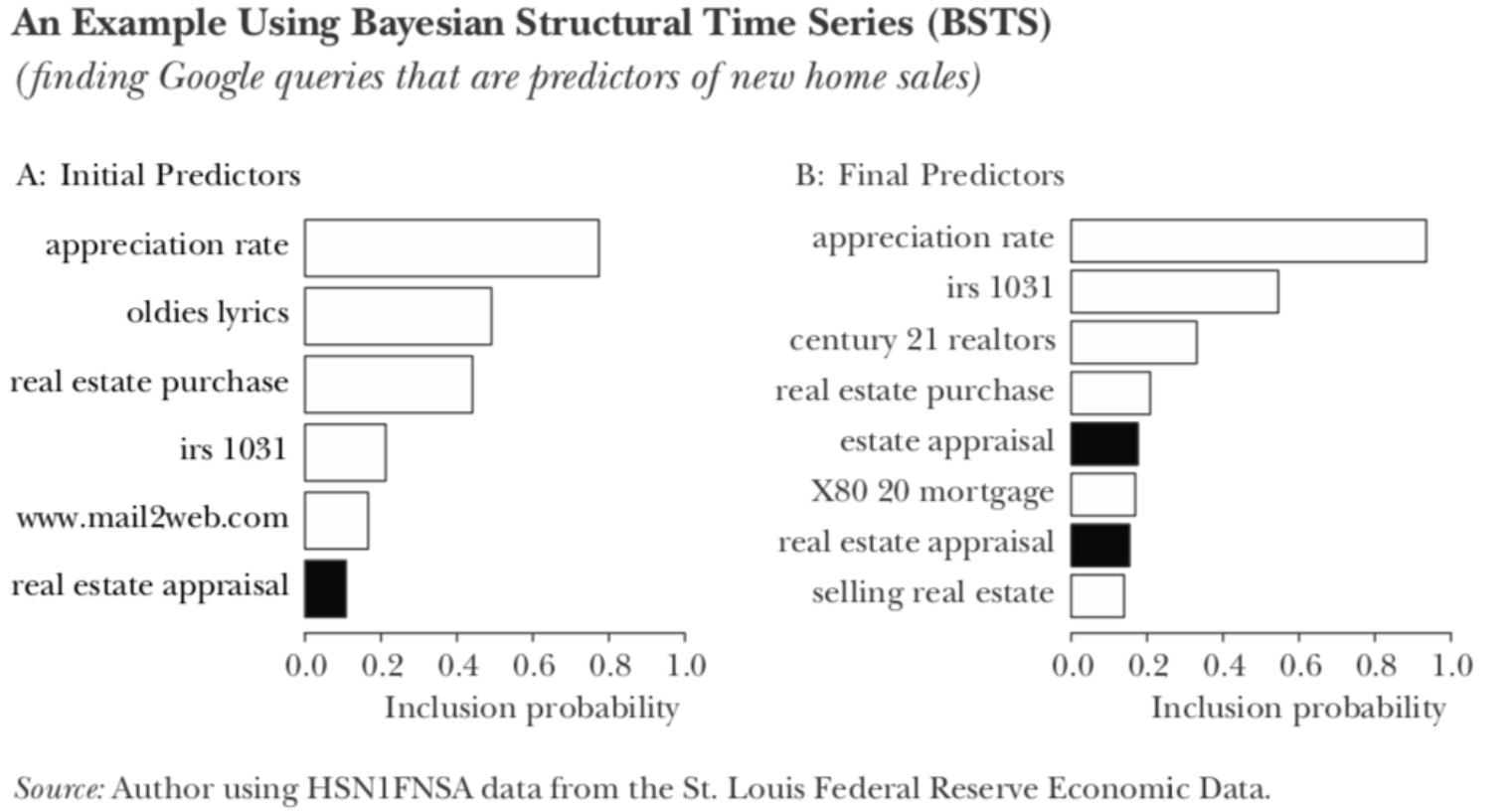
- Data has signal and noise.
- More "expressive" functions
  - Capture the signal better
  - But also pick up the noise
- Prediction algorithms allow tuning of parameters so that the signal/noise ratio is as desired

Example: Prediction gone wrong



Source: Lazer et al. 2014.

Example: Automated selection



Panel A shows uncurated Google Correlate results to predict new homes sold in the US. Choices adjusted by hand in Panel B

Source: Varian, 2014.

Why can predictions go wrong?

- Let's go back to the simple model, and note that it most often is a vector valued (see e.g. Facebook example below)

$$Y_i = f(\boldsymbol{X}_i) + \epsilon_i$$

- Machine learning algorithms choose what parameters of $\boldsymbol{X}_i$ are used, automatically and based on past data
- With big data, manual analysis of data would be impossible/too costly/too slow
- In complex environments with big data, no one knows why a certain set of parameters gets chosen

Experiments and A/B testing

- Randomized experiments used to develop new innovations
- For example, take a random sample of users who are shown a version B of the website while the the rest use version A
- Comparing results (e.g. clicks, purchases) between A and B can *predict* which version is better
- These are very similar to controlled experiments used in economics, medicine etc., just the objective is different
- For a website it may be enough to see which version works better (for now), but for other decisions it can be crucial to understand what are the reasons for the differences

Simple solutions: Change one parameter at a time

- Question: What price to charge customers?
- Direct variation of prices can be problematic, instead indirect means are used to extract data (see Figure).
- Such testing, done offline as well, can provide an understanding on how demand changes when prices change.
- Online market places differ from traditional sales in the speed and reach of testing.

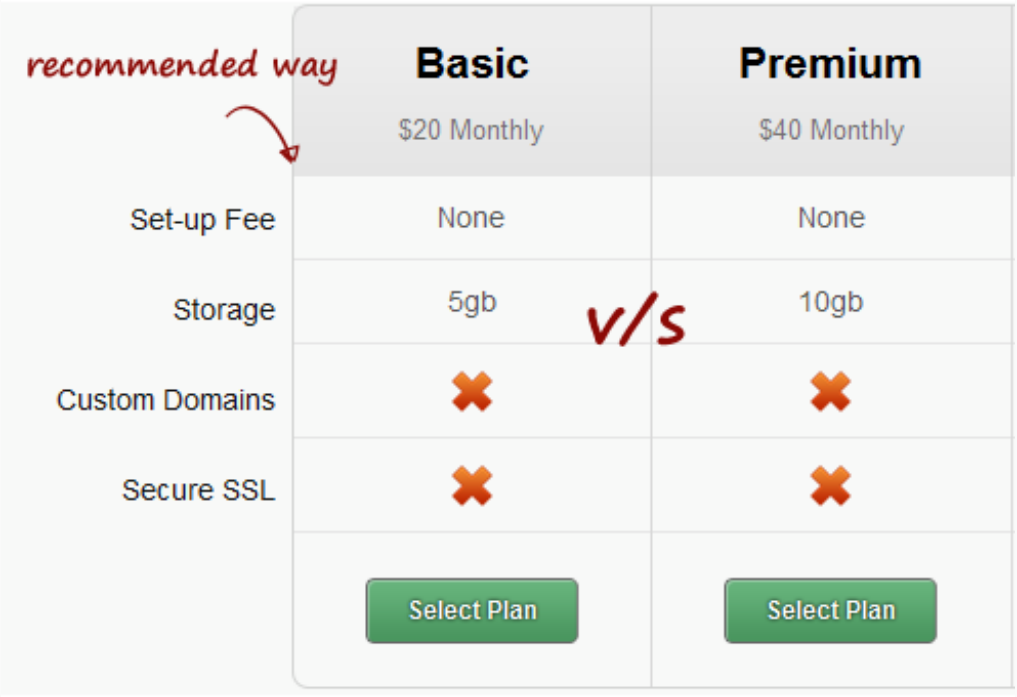


Figure: VWO.

More complex example: A/B testing at Microsoft

- Microsoft Bing has a dedicated platform to run A/B tests
- Nearly costless to test: thousands of new ideas per year
- Scarcity of good ideas, and surprisingly, data as well
- Innovations are typically A/B tested for 1 week, with an average of about 20 million users
- Large sample sizes needed as most innovations lead to small improvements

Source: Azevado et al. (2020).

Further example: Causal inference with prediction

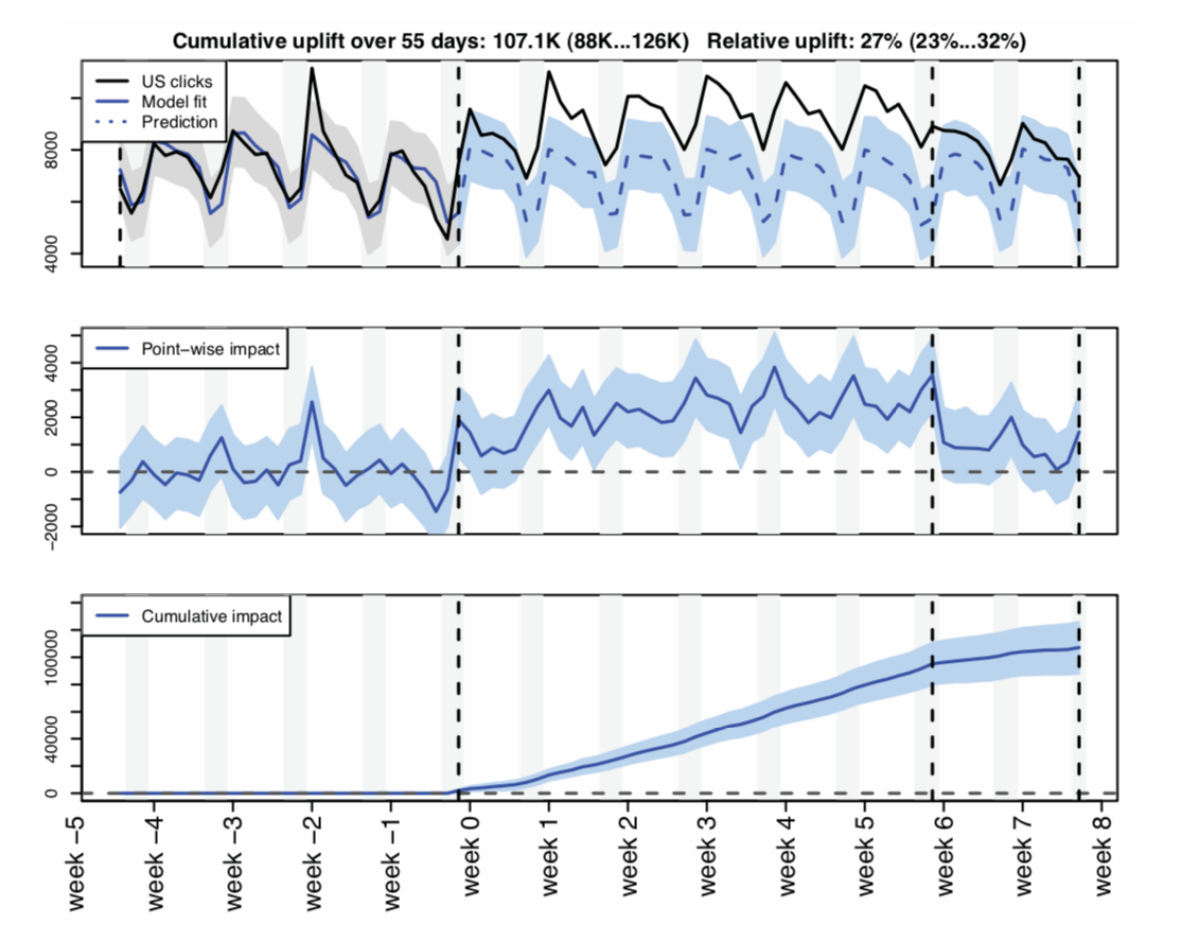
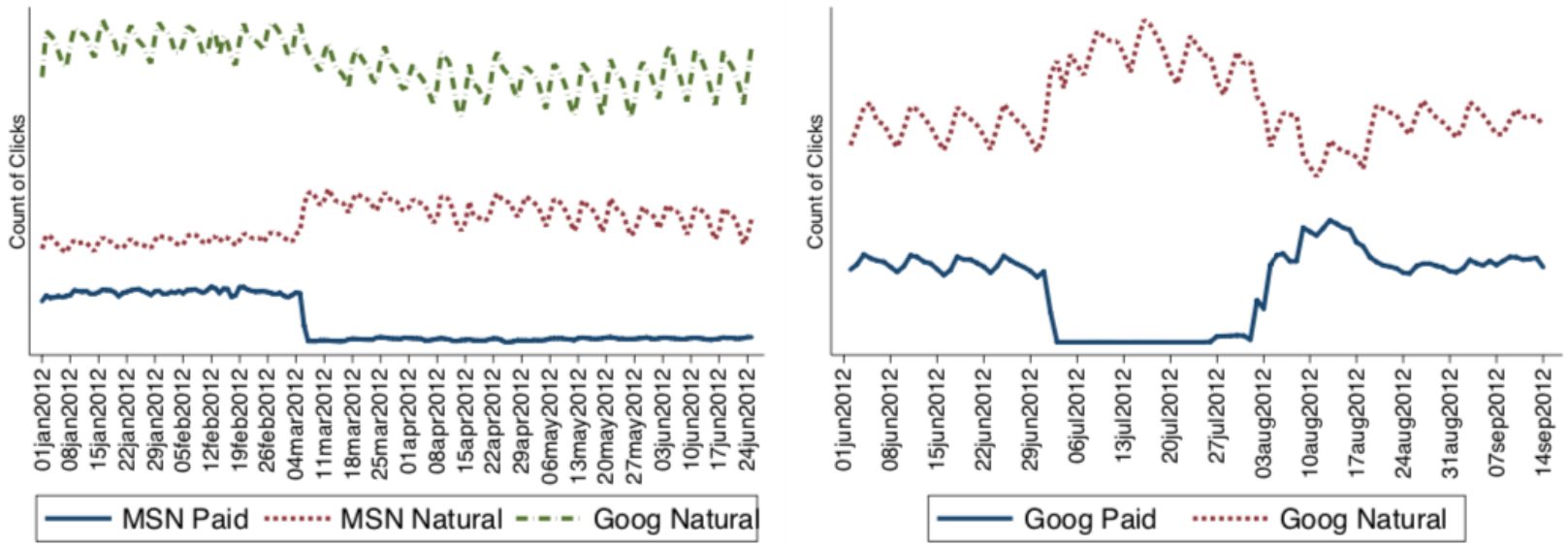


Figure. Actual clicks vs. predicted clicks and 95 % confidence interval.

Varian (2014).

Care needed in interpretation

Figure 2: Brand Keyword Click Substitution



(a) MSN Test **(b) Google Test**

Note: MSN and Google click traffic is shown for two events where paid search was suspended (Left) and suspended and resumed (Right).

Figure. Regular search is a substitute for eBay brand ads. One explanation: selection of consumer who want to go to eBay.

Blake et al. (2014).

Takeaways from today

- Firms extract data because it is of value to them
- Collecting data online easier than ever
- Prediction and causal inference aim for different goals
  - Firms mostly interested in prediction
  - Care needed to understand what are the best tools

Materials for this week

Reading assignment 1:

- Athey, Susan and Michael Luca (2019) "Economists (and Economics) in Tech Companies", Journal of Economic Perspectives. Read the whole article.
- Varian, H. (2012) "Revealed Preferences and its Applications", Economic Journal. Read the Introduction and Section 2 for now.

Online resources (make sure you know these before you take Exercise 1):

- Make sure you know the basics of consumer choice: e.g. mru.org: Consumer Choice and/or www.core-econ.org 3.2-3.5, 3.7.1.
- and supply–demand equilibrium e.g. mru.org: Supply, Demand, and Equilibrium and/or www.core-econ.org 8.1, 8.2.

Next time

Competition

- Perfect competition
- Monopoly
- Oligopoly