Advertising Measurement

Template Test — UCSD MGTA 451

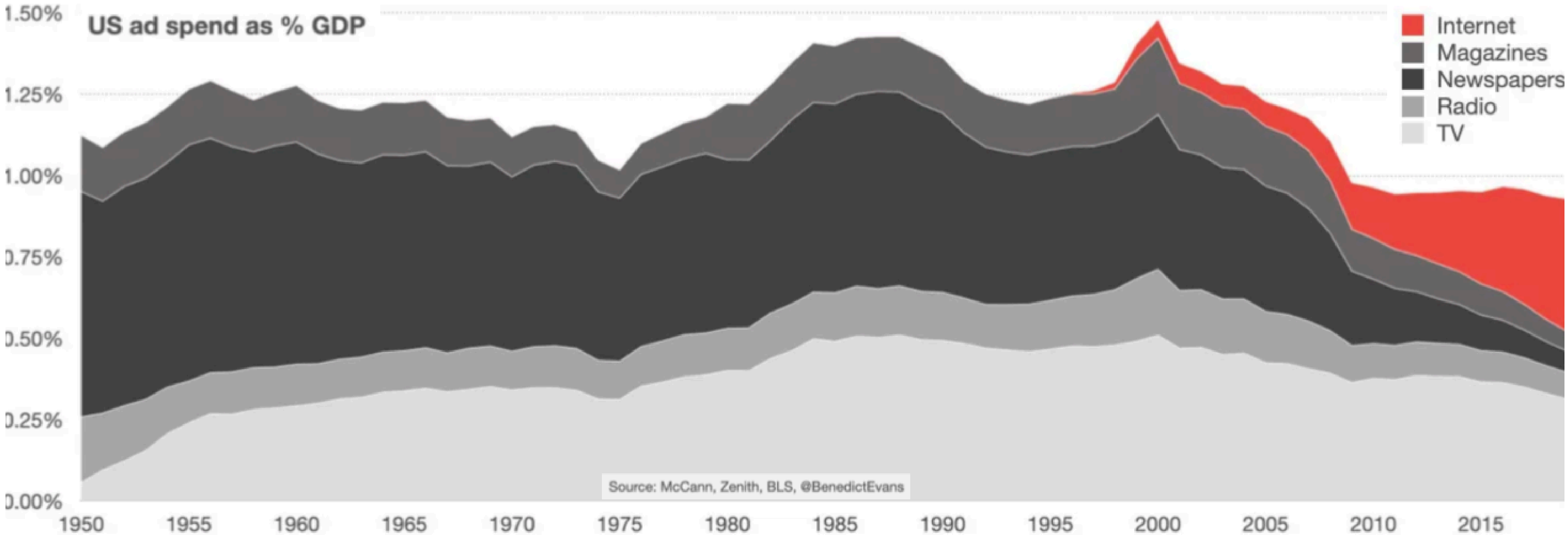
Kenneth C. Wilbur

UC San Diego

Agenda

- Advertising Importance
- Causality
- Fundamental Problem of Causal Inference
- Advertising Measurement
- Correlational Advertising Measurement
- Causal Advertising Measurement
- Industry practices
- Marketing Mix Models
- Career considerations

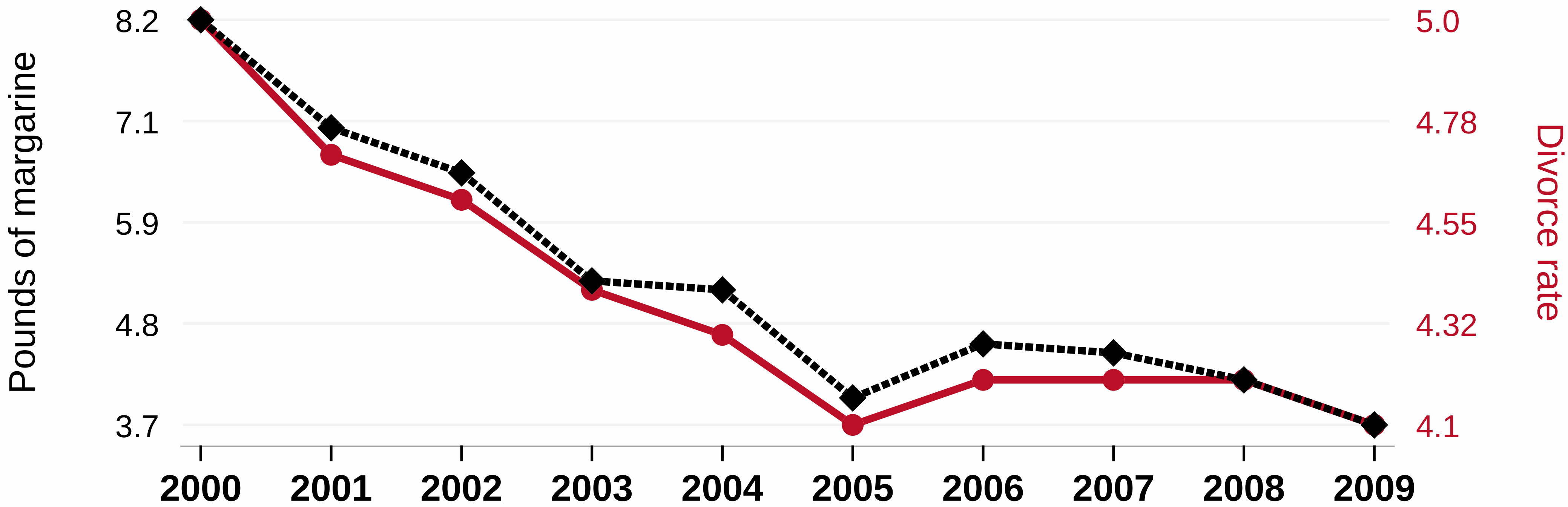
Publisher revenue since 1950



Per capita consumption of margarine

correlates with

The divorce rate in Maine



Per capita consumption of margarine in the United States · Source: US Department of Agriculture

The divorce rate in Maine · Source: CDC National Vital Statistics

2000-2009, r=0.993, r²=0.985, p<0.01 · tylervigen.com/spurious/correlation/5920

Classic misleading correlations

- “Lucky socks” and sports wins

```
- Post hoc fallacy [1] (precedence indicates causality AKA superstition)
```

- Commuters carrying umbrellas and rain

```
- Forward-looking behavior
```

- Kids receiving tutoring and grades

```
- Reverse causality / selection bias
```

- Ice cream sales and drowning deaths

```
- Unobserved confounds
```

- Correlations are measurable & usually predictive, but hard to interpret causally

```
- Correlation-based beliefs are hard to disprove and therefore sticky
- Correlations that reinforce logical theories are especially sticky
- Correlation-based beliefs may or may not reflect causal relationships
```

“Revenue too high alert”

Why care?

- We want to maximize profits $\Pi = \Sigma_i \pi_i(Y_i(T_i), T_i)$
- Suppose $Y_i = 1$ contributes to revenue; then $\frac{\partial \pi_i}{\partial Y_i} > 0$
- Suppose $T_i = 1$ has a known cost, so $\frac{\partial \pi_i}{\partial T_i} < 0$
- Effect of $T_i = 1$ on π_i is $\frac{d\pi_i}{dT_i} = \frac{\partial \pi_i}{\partial Y_i}\frac{\partial Y_i}{\partial T_i} + \frac{\partial \pi_i}{\partial T_i}$
- We have to know $\frac{\partial Y_i}{\partial T_i}$ to optimize T_i assignments

    - Called the "treatment effect" (TE)

- Profits may decrease if we misallocate T_i

Fundamental Problem of Causal Inference

- We can only observe **either** $Y_i(T_i = 1)$ **or** $Y_i(T_i = 0)$, but not both, for each person i

```
- The case we don't observe is called the "counterfactual"
```

- This is a missing-data problem that we cannot resolve. We only have one reality

```
- A major reason we build models is to compensate for missing data
```

3. Multi-Touch Attribution (MTA)

Get individual-level data on every touchpoint for every purchaser

- Includes earned media (PR, reviews, organic social), owned media (website, content marketing, email) & paid media (<--ads; also, paid influencer & affiliate)
- Often sourced from third parties

Choose a rule to attribute purchases to touchpoints

- Single-touch rules: Last-touch, first-touch
- Multi-touch rules: Fractional credit, Shapley

Historically, Last-touch was popular

MTA algorithm searches for touchpoint parameters that best-fit the conversion data given the rule

- Credit then informs future budget allocations
- MTA is designed to maximize attributions
- MTA assumes advertising is the *sole* driver of conversions

MTA is mostly dead due to privacy and platform reporting changes

- Governments, platforms, browsers, OS have all restricted MTA input data for privacy
- Some advertisers' MTA lives on due to inertia, despite signal loss
- Large platforms offer MTA results within the platform

U.S. v Google (2024, search case)

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA et al.,	)	
Plaintiffs,	)	
v.	)	**Case No. 20-cv-3010 (APM)**
GOOGLE LLC,	)	
Defendant.	)	

263. When it made pricing changes, Google took care to avoid blowback from advertisers. For instance, records show that Google had concerns about the impact of transparency on their efforts to increase prices. *See* UPX507 at .015 ("Worry that if we tell advertisers they will be impacted, they will attempt to game us and convince us to abandon the experiment. . . . But, if influence our decision at all)."); UPX519 at .003 ("A sudden step function might create adverse reaction.").

264. Google therefore endeavored to raise prices incrementally, so that advertisers would view price increases as within the ordinary price fluctuations, or "noise," generated by the auctions. *See, e.g.*, UPX507 at .023 (describing a 10% CPC increase as "safe" because it is "within usual WoW noise"); UPX519 at .003 (acknowledging that advertisers would notice a 15% price increase, but "this change is to [be] put in perspective with CPC noise," that is, "50% of advertisers seeing 10%+ WoW CPC changes"); *id.* (comment stating that 15% is "probably an acceptable level of change (from a perception point of view) because these are magnitudes of fluctuations they are used to see[ing]").

265. With respect to format pricing, one Google document states: "A progressive ramp up leaves time to internalize prices and adjust bids appropriately[.]" UPX519 at .003; UPX509 at 870 (stating that "[i]ncremental launches and monitoring should help us manage" the risk that price increases would lead advertisers to "lower[] their bids or modify[] other settings . . . to get back to a given ROI, leading to less revenue for Google than the initial impact hinted to"). Similarly, in 2020, Google raised prices on navigational queries using multiple knobs and recognized that it was "[o]bviously a very large change that we don't intend to roll out at once," instead planning a "[s]low 18 months rollout" to "[l]eave[] time for advertiser[s] to respond rationally[.]" UPX503 at 034; *id.* at 038 ("A slow roll ensures we don't shock the system, gives time for advertisers to respond and us to monitor changes and stop early if needed."); *see also, e.g.*, UPX505 at 312 (prior to implementing squashing, concluding that "[a]dvertisers should perceive AdWords as a consistent system, and not be subject to constant large impacts due to Google changes," in part to "improve[] advertiser stickiness"); UPX506 at .018 (Momiji slide deck: "Unlikely that advertisers will notice by themselves and respond. However, a bad press cycle could put us in jeopardy.").

266. Google's incremental pricing approach was successful. In 2018 and 2019, Google conducted ROI Perception Interviews, which raised no red flags about advertisers' attitudes as to ad spending on Google. *See generally* DX187; DX119. While advertisers could tell that prices were increasing, they did not understand those changes to be Google's fault. Google's studies revealed that advertisers facing CPC changes "dominantly attribute[d] these shifts to themselves, competition[,] and seasonality (85%)—not Google." UPX1054 at 061; *see also* UPX737 at 464 ("They often attribute these changes to things in the world or what they've done, not just things happening on the backend[.]").

CONCLUSION

For the foregoing reasons, the court concludes that Google has violated Section 2 of the Sherman Act by maintaining its monopoly in two product markets in the United States—general search services and general text advertising—through its exclusive distribution agreements. The court thus holds that Google is liable as to Counts I and III of the U.S. Plaintiffs' Amended Complaint, Am. Compl. ¶¶ 173–179, 187–193. To the extent that Counts I and III of the Plaintiff States' Complaint are co-extensive with the U.S. Plaintiffs' Counts I and III, the court finds Google liable. *Colorado* Compl. ¶¶ 212–218, 226–232.

Amit P. Mehta
United States District Court

Ad Experiments: Common Designs

1. Randomly assign ads to customer groups on a platform; measure sales in each group

```
- Pros: AB testing is easy to understand, rules out alternate explanations
- Cons: Can we trust the platform's "black box"? Will we get the data and all available insights?
```

2. Randomize messages within a campaign. Mine competitor messages in **ad libraries** for ideas

```
  - Often a great place to start
```

3. Randomize budget across times & places (“Geo tests”)
4. Randomize bids and/or consumer targeting criteria
5. Randomize budget over platforms, publishers, behavioral targets, contexts

```
  - Experimental design describes how we create data to enable treatment/control comparisons. Experimental data are amenable to any
number of models or statistical analyses.
  - Causal identification is a property of the data, not the model
```

Muy importante

Before you kick off your test ...

- Run A:A test before your first A:B test. Validate the infrastructure before you rely on the result. A:A test can fail for numerous reasons

- Can we agree on the opportunity cost of the experiment? "Priors"

- How will we act on the (uncertain) findings? Have to decide before we design. We don't want "science fair projects"

- Simple example: Suppose we estimate iROAS at 1.5 with c.i. [1.45, 1.55]. Or, suppose we estimate RoAS at 1.5 with c.i. [-1.1, 4.1]. What actions would follow each?