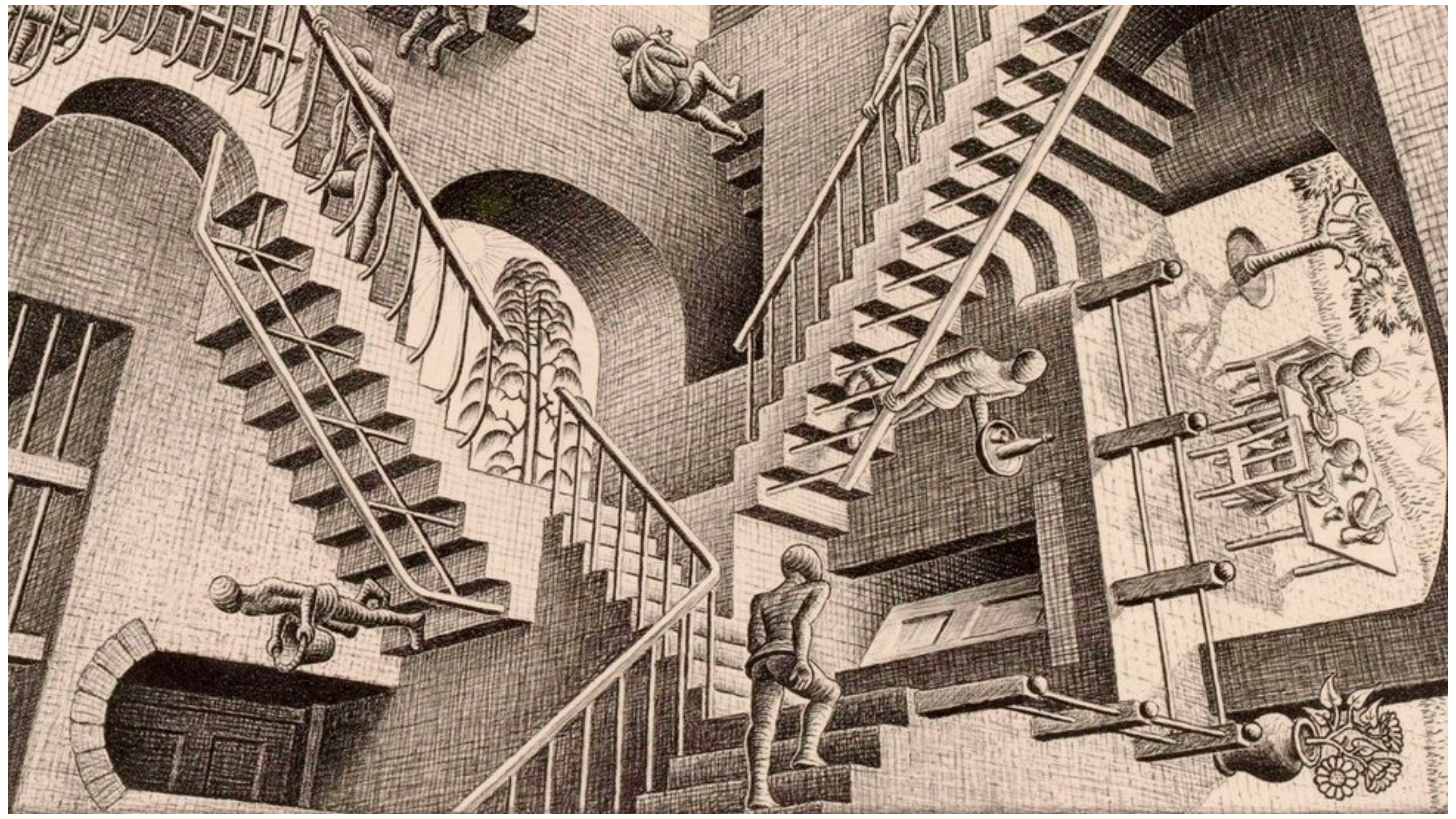
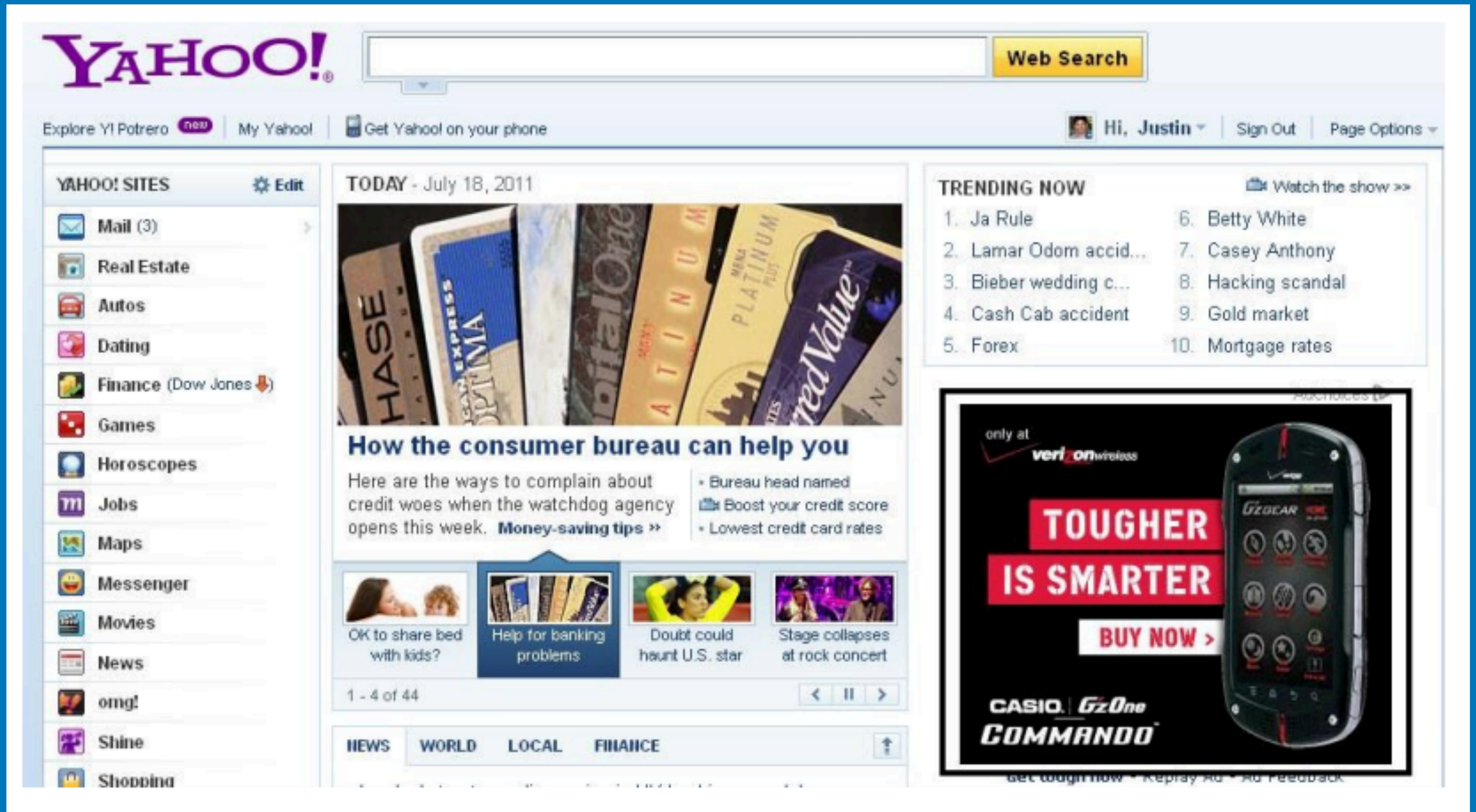
Business Experimentation and Causal Methods

Prof. Fradkin
Topic: The Difficulty of Measuring ROI



On the Near Impossibility of Measuring Returns to Advertising

Lewis and Rao (2015)



Above Paper

- 25 large-scale online display advertising field experiments.
- $2.8 million in expenditure.
- Over 1MM observations for most of the experiments.
- Outcome: Profits and revenue

What they measured

- $\hat{ATE}$ = \$0.35 (effect on revenue)
- Cost per exposed person = .14
- SD of Revenue = \$75
- Margin = .5

Lift: Another Common Measure of Ad Effects

- Sometimes we might state this in terms of % change:

$$\widehat{\text{lift}} = 100 \times \frac{\widehat{\text{ATE}}}{\bar{Y}(0)}$$

- This doesn't say anything about change in mean relative to variance.
- If we assume average revenue of \$8 and ATE is \$.35 then lift is 4.4%.

Signal

$0.35

ad effect

Noise

$75

Standard deviation of sales

Illustration by Garrett Johnson

How many observations to detect this?

```python
from statsmodels.stats.power import TTestIndPower

# Calculate the power
n = power_analysis.solve_power(effect_size=.35/75, power = 0.8, alpha=alpha, ratio=1, alternative='two-sided')

print(f"Necessary sample size: {np.ceil(n)}")
```

✓ 0.0s

```
Necessary sample size: 720815.0
```

But what about positive ROI?

- Our null hypothesis should instead be ROI = 0, since that is the threshold for whether it makes sense to advertise.

- Cost per impression is .14, margin is .5.

- Estimate of ROI is (.35*.5 - .14)/.14 = .25
(effect on sales * margin - cost) / cost)
(assume no uncertainty in cost (often a bad assumption))

- SD of ROI: 75*.5/.14 = 267.8571.

```
n = power_analysis.solve_power(effect_size = .25/267.8571,
                               nobs1 = None, ratio = 1, alpha = .05, power = 0.8, alternative='two-sided')

print(f"Necessary sample size: {np.ceil(n)}")
```

✓ 0.0s

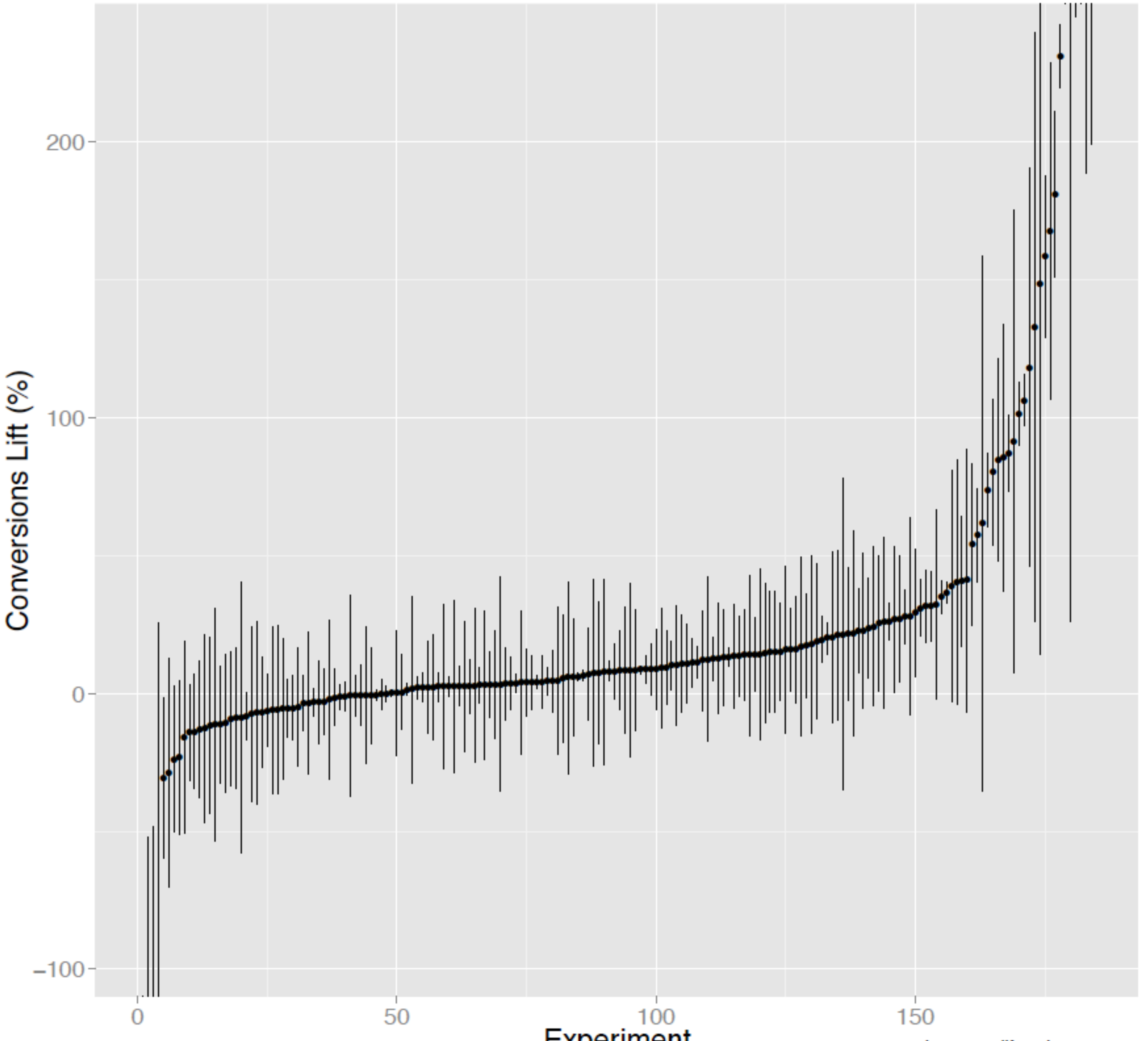
```
Necessary sample size: 18020339.0
```

Over 18 Million Obs!

Most display ad campaigns don’t work, but some do.

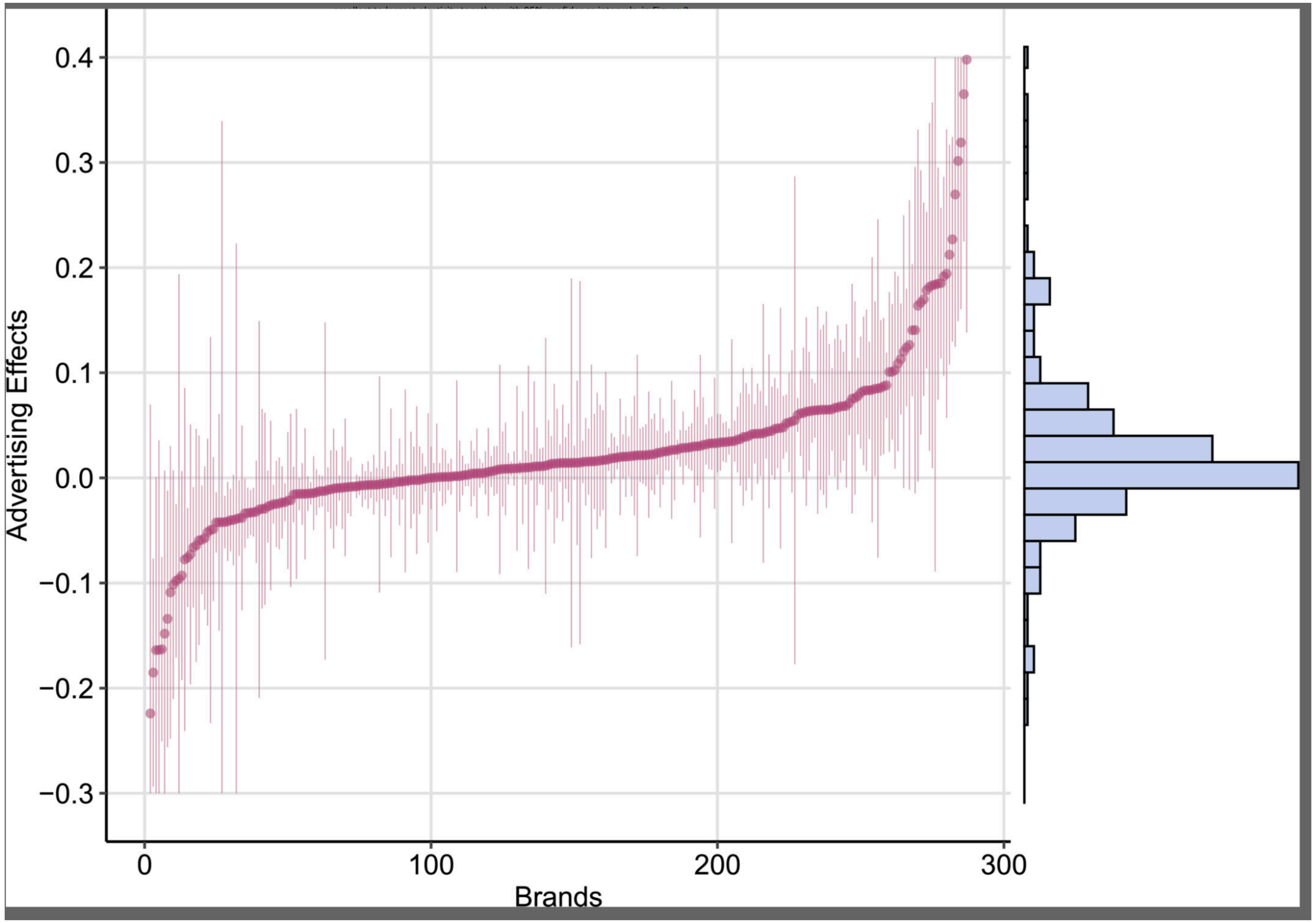
Point is the estimated lift on conversions.

Each line is 95% confidence interval for an experiment.



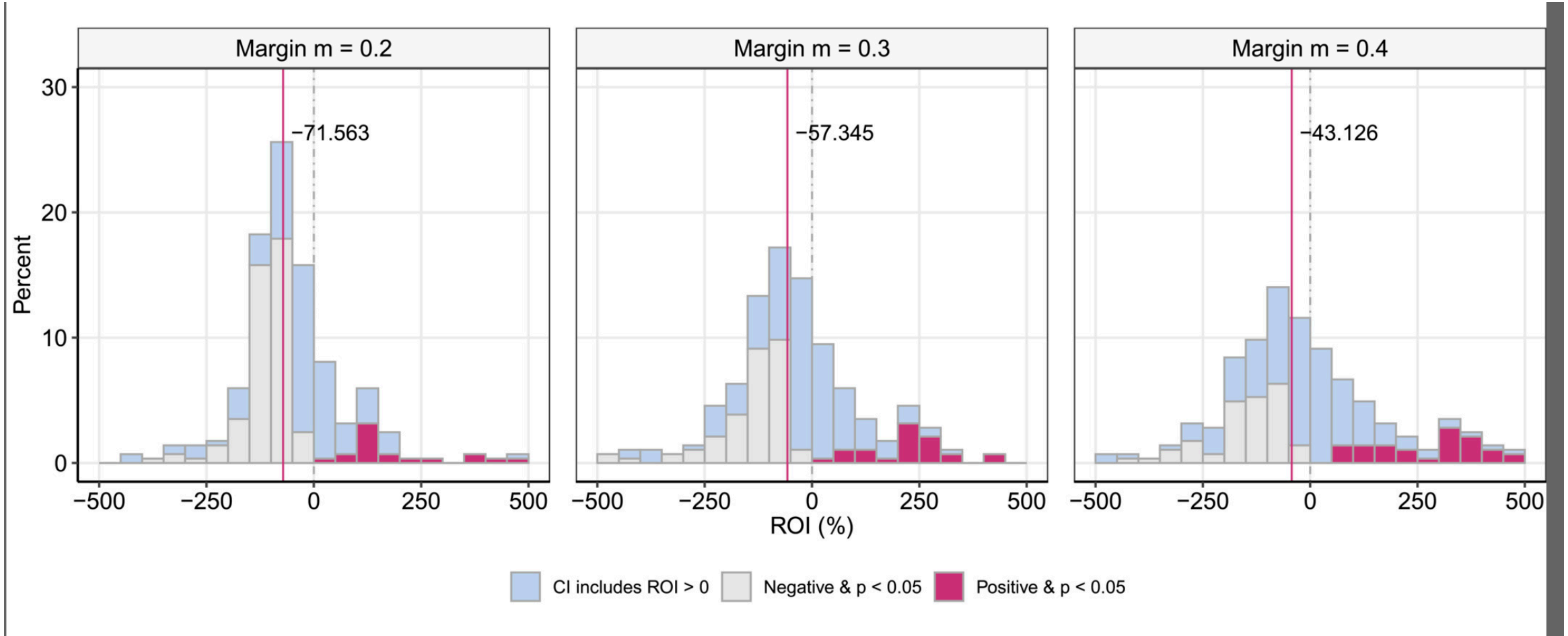
Johnson, Garrett and Lewis, Randall A. and Nubbemeyer, Elmar, The Online Display Ad Effectiveness Funnel & Carryover: Lessons from 432 Field Experiments (October 1, 2017). Available at SSRN: https://ssrn.com/abstract=2701578 or http://dx.doi.org/10.2139/ssrn.2701578

Similar for TV ads



Source: Shapiro, Bradley T., Günter J. Hitsch, and Anna E. Tuchman. "TV advertising effectiveness and profitability: Generalizable results from 288 brands." *Econometrica* 89.4 (2021): 1855-1879.

Similar for TV ads



(b) ROI of all Observed Advertising

Source: Shapiro, Bradley T., Günter J. Hitsch, and Anna E. Tuchman. "TV advertising effectiveness and profitability: Generalizable results from 288 brands." *Econometrica* 89.4 (2021): 1855-1879.

When is ad experimentation likely to … ?

Detect Effects (high power)

- Large N
- Expecting big effects, spending more per person
- People purchase regularly or frequently.
- New products. (Low variance of revenue in control group since they don’t know about the product)
- Target people who we have bigger effects for.

Not Detect Effects (low power)

- Other ad campaigns going on in control, or other unobserved factors driving control sales
- Control group is being affected by the treatment.
- Infrequently purchased products, big tickets items. (High variance of outcome)

Conclusion

Lots of money spent in advertising.

Typical effects are small, and need experiments with very large sample sizes to detect positive ROI.