

Estimating the Value of Offsite Data to Advertisers: Evidence from Meta

Nils Wernerfelt Anna Tuchman Bradley T. Shapiro Robert Moakler

November 2, 2023

Offsite data is widely used but faces an uncertain future

- Online activity is frequently tracked and shared across applications to help target digital advertising.
 - E.g., website browsing behavior, online purchases
- Major ad platforms often use a form of web pixel to track and share such “offsite” data.
 - Meta, Google, Twitter, TikTok, Snap, etc.
- Regulation and product changes increasingly threaten the ability of advertisers to use this data (appeal to consumer privacy).
- Holistic evaluation requires understanding the value of offsite data for advertising effectiveness.

We estimate ad effectiveness with and without such data

- We do two things for a large sample of advertisers:
 - 1) Take live campaigns that use offsite data and randomly hold users out from exposure.
 - 2) Adjust a small fraction of traffic from those campaigns to use onsite data instead and again randomly hold users out.
- Within campaigns, what is the cost per incremental customer at baseline with offsite data and how does it shift without such data?

Two Main Contributions

- 1) Large scale study of the effectiveness of digital advertising on purchasing behavior on a major platform
 - 70k+ advertisers in our sample, minimal selection
 - Flexibly estimate entire distribution of effects (Efron, 2016)
- 2) Generalizable evidence on the effect of losing offsite data on advertising effectiveness

Caveats and Cautions Upfront

- Partial equilibrium
 - In GE, ad prices may adjust, advertisers substitute off platform, etc.
 - Platforms might innovate targeting technology in the long run
- Cannot make any statements about social welfare
 - E.g., we don't measure value of privacy to consumers
- Other platforms, other ways advertisers use offsite data

Outline

- Background
- Experimental Design
- Sample
- Main Results
- Additional Results
- Conclude

Background: Digital Advertising Context

Budget & schedule

Budget ⓘ

Daily Budget USD

You'll spend up to \$312.50 on some days, and less on others. You'll spend an average of \$250.00 per day and no more than \$1,750.00 per calendar week. [Learn more](#)

Schedule ⓘ

Start date

Oct 27, 2022

Pacific Time

End - Optional

Set an end date

Choose a campaign objective

- Awareness
- Traffic
- Engagement
- Leads
- App promotion
- Sales



Sales

Find people likely to purchase your product or service.

Good for:

- Conversions ⓘ
- Catalog sales ⓘ
- Messenger and WhatsApp ⓘ
- Calls ⓘ

Audience

Define who you want to see your ads. [Learn more](#)

[Create new audience](#) Use saved audience ▾

Custom audiences Create new ▾

[Exclude](#)

Locations

Location:

- United States

Age

20 - 30

Gender

All genders

Detailed targeting

Include people who match ⓘ

[Interests > Business and industry](#)

Marketing

Background: Intuition behind optimization

Page Likes Example

- Suppose you want to generate likes on your business page
- One approach for ad delivery: uniformly distribute amongst target audience
- But – can do better incorporating empirical response data of objective
- Can use that information to train a model to predict $P(\text{like page})$, and then show the ad only to people with high predicted value

Background: Offsite data, pixels

- Can use same machinery with other outcomes – notably, *purchases*
- This is where pixels enter:
 - Install pixel on website; fires when someone makes a purchase event
 - Allows purchases to be a left hand side variable
- Delivery optimization is arguably a major upside of digital advertising
- **Note:** may not be incremental, a point we take seriously in our design.

Not just a Meta product

- Google
- Twitter
- Pinterest
- TikTok
- Snapchat
- LinkedIn
- ...

The image displays three screenshots of advertising campaign creation interfaces:

- Google Ads Help:** A page titled "About Smart Bidding" with a search bar and navigation links. The main content explains Smart Bidding as a machine learning strategy to optimize conversions or conversion value. A sidebar lists links for "About automated bidding", "Your guide to Smart Bidding", "About Smart Bidding", and "Create a portfolio bid strategy".
- Twitter Business:** A dark-themed banner for "Create a website traffic campaign" with a background image of a car's engine. Navigation links for "Business", "Basics", "Advertising", "Targeting", "Reporting", "Resources and guides", "Ads Help Center", "Sign in", and "Start a campaign" are visible.
- TikTok Business Help Center:** A page titled "What's the Website Conversions Objective" with a navigation bar for "Basics", "Ads", "Management", "Measurement", "Billing", and "Resources". The main content explains the website conversions objective. A sidebar includes "Create a Campaign" and "What's the Website Conversions Objective". A yellow banner below reads "Business Help Center" with a search bar and "How can we help?".

Background: Recap and our focus case

- We focus on a primary use case of offsite data: *offsite conversion optimization*.
 - Relies on offsite data to generate LHS variable in a prediction problem
- We take large sample of advertisers who are optimizing for purchases, measure how effective their ads are vs. optimizing for onsite outcome
 - **In short:** Taking X's as fixed, varying Y's; if advertiser optimizes today for one vs. other

Background: Our Counterfactual

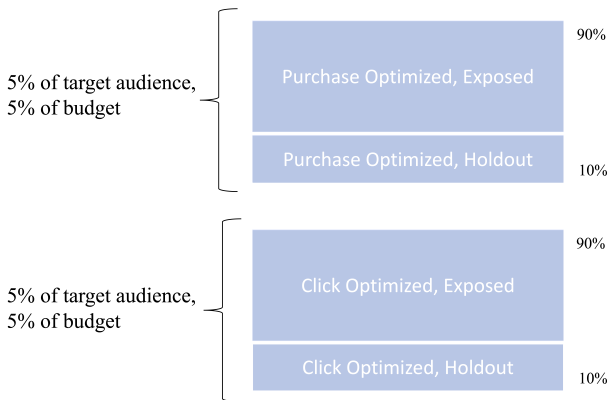
- If advertisers cannot use offsite conversation optimization, what would they do instead?
- Our counterfactual: *click optimization*
 - Show ads to users who are predicted to click on the ad
 - Lowest outcome in conversion funnel observed onsite
 - Also a popular optimization objective

Outline

- Background
- **Experimental Design**
- Sample
- Main Results
- Additional Results
- Conclude

Experimental Design

- One experiment: 10% of traffic from all ads optimizing for a purchase event on a pixel from our advertisers.
- Holdout: focal ad withheld and second place ad sent (standard 'lift' infrastructure)



Outline

- Background
- Experimental Design
- **Sample**
- Main Results
- Additional Results
- Conclude

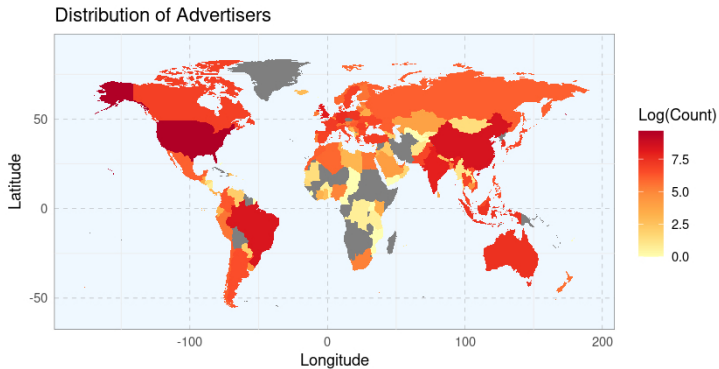
Sample: Near universe of relevant advertisers

- Sent opt-out notice to near universe of advertisers who used offsite conversion optimization in three months prior to experiment (31 languages)
- High percentage did not opt out (94%).
- After cleaning, left with 70,909 experiments
 - Note: large number are using 'incorrectly'

The screenshot shows a mobile application interface for managing advertisements. At the top, the time is 4:12 and there are status icons for signal, Wi-Fi, and battery. Below the time is a navigation bar with a back arrow and the word "Ads". Underneath are two tabs: "Create" and "Manage", with "Manage" being the active tab. A modal dialog box is displayed over the "Manage" tab, titled "Participate in Advertiser Study". The dialog contains the following text: "You've been randomly selected to participate in a research study on ad effectiveness. The goal of this study is to measure how effective different advertising practices are at driving sales, thereby helping us to invest in them the future. As part of that, we will randomly deliver a small fraction of your ads over the next three weeks according to a different optimization goal or lookalike model if you selected those options. We expect this will be within the range of the normal variation we'd see in the delivery system, but if you would like to opt out, you are welcome to do so by clicking the below button." Below the text is a blue button labeled "Opt out". Below the dialog is a section titled "Advertising Summary" with the text: "When you create ads, you'll get an overview of how they're performing." Underneath this is a dropdown menu labeled "Currently viewing" with "Last 60 days" selected. At the bottom of the screen are three buttons: "People Reached", "Post Engagements", and "Link Clicks". The bottom right corner shows navigation icons and the page number "16/26".

Sample: Spans geographies, verticals

- Advertisers from 161 countries
 - US (22%), China (7%), Brazil (6%), India (4%)
- E-commerce (44%), Retail (19%), CPG (12%)
 - Within E-commerce, mostly apparel and household goods

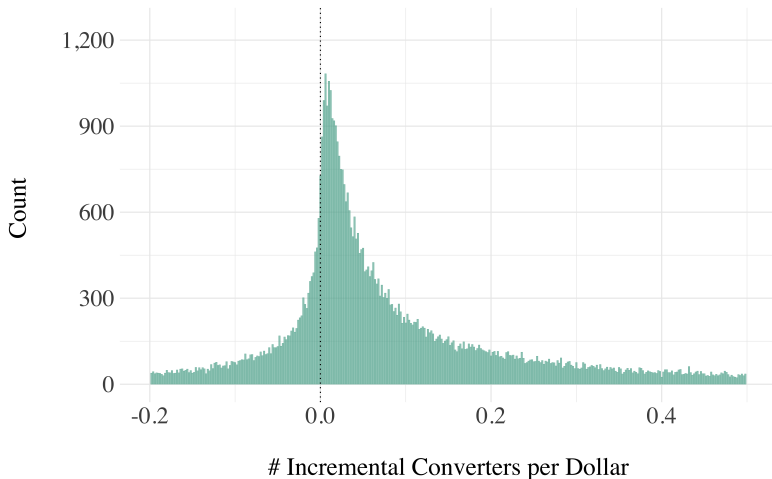


Outline

- Background
- Experimental Design
- Sample
- **Main Results**
- Additional Results
- Conclude

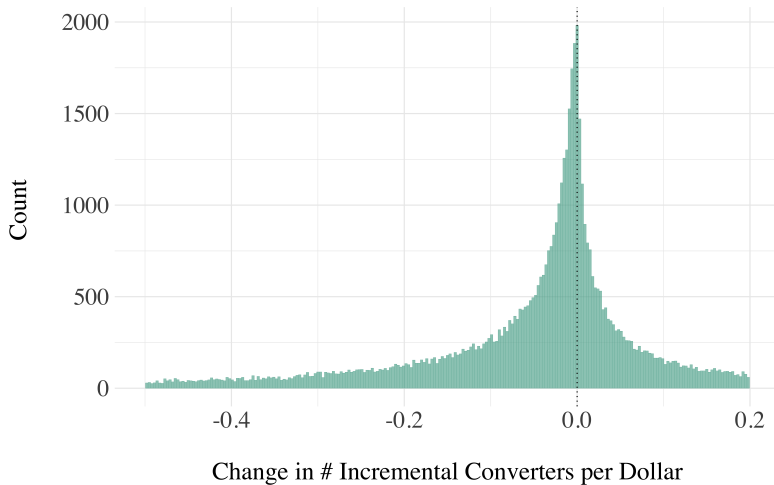
Empirical Distribution of Treatment Effects, Baseline

Histogram of # Incremental Converters per Dollar, Baseline



Change In Effectiveness: Click - Purchase Optimized

Histogram of the Change in # Incremental Converters per Dollar



Detailed Results on Estimated Distributions

Table: Summary statistics of estimated distributions (Efron, 2016).

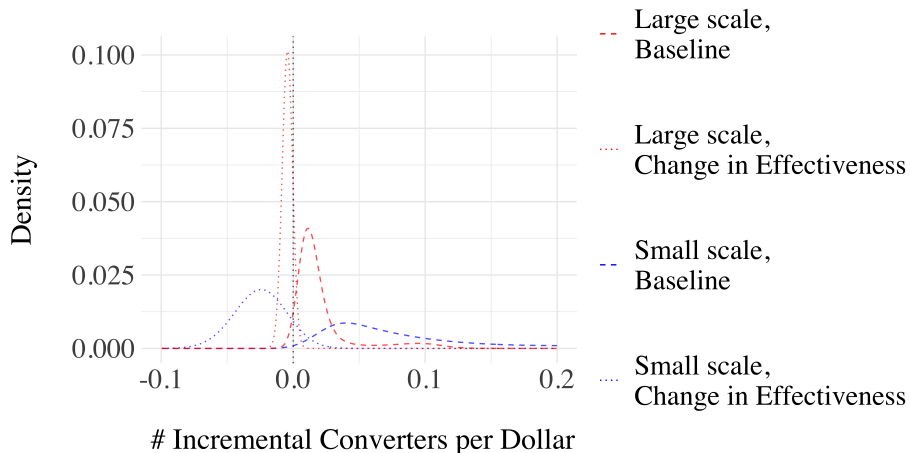
	10th	25th	50th	75th	90th	Mean
Baseline Effectiveness						
# Incremental Converters per \$1,000	4.8	11.8	23.8	70.8	189.8	90.4
Change in Effectiveness						
# Fewer Incremental Converters per \$1,000	-12.2	-9.2	-6.2	-3.2	-0.2	-7.0

⇒ At the median estimates, the cost per incremental converter increases from \$42.04 to \$56.77, roughly a 35% increase.

Outline

- Background
- Experimental Design
- Sample
- Main Results
- **Additional Results**
- Conclude

“Small” vs. “Large” Scale Businesses



Long Term Effects

- We also follow up on users 6 months after our experiment and analyze their purchasing behavior, knowing the initial random assignment
- Two reasons long term effects are important:
 - **Firm side:** Firms care about lifetime value of customers
 - **User side:** Are ads helping or hurting consumers? (suggestive)
 - Look at revealed preference around long term buying behavior under each kind of advertising
- We find ads delivered with offsite data generate more long term customers per dollar than ads delivered without.
 - Cost per incremental 'long term' customer 16% higher without offsite data

Further Implications

- Our experiment suggests offsite data substantially improves ad effectiveness on Meta. What are further implications for different parties?
- **Advertisers**
 - Willingness-to-pay for offsite purchase data
 - Potential gains from trade from compensating users for data
- **Platforms**
 - Value of bringing offsite data onsite (e.g., Shops)
 - Invest in privacy enhancing technologies
- **Extending beyond our estimates**
 - Potential competitive implications (product markets, digital advertising)

Conclusion

- We focus on estimating the value of offsite data to advertisers on Meta. Leverage a representative sample of 70k+ advertisers.
 - These data are believed to be important for a large share of digital advertising, and current gap in literature around their value.
- We find evidence ad effectiveness would be substantially hampered by loss of this data (35% increase in costs for median advertiser, median loss.)
- We find evidence that losing offsite data hurts smaller scale advertisers more and increases costs for attracting long term customers.

Estimated Distribution of Ad Effectiveness

